



---

**Statistical Policy  
Working Paper 15**

**Quality in Establishment Surveys**

---

**Prepared by  
Subcommittee on Measurement of Quality in Establishment Surveys  
Federal Committee on Statistical Methodology**

**Statistical Policy Office  
Office of Information and Regulatory Affairs  
Office of Management and Budget**

**July 1988**

DDP 210  
CIT 170

**MEMBERS OF THE FEDERAL COMMITTEE ON  
STATISTICAL METHODOLOGY**

**(July 1988)**

**Maria E. Gonzalez (Chair)**  
**Office of Management and Budget**

**Yvonne M. Bishop**  
**Energy Information**  
**Administration**

**Warren L. Buckler**  
**Social Security Administration**

**Charles E. Caudill**  
**National Agricultural**  
**Statistics Service**

**Edwin J. Coleman**  
**Bureau of Economic Analysis**

**Charles D. Cowan**  
**National Center for Education**  
**Statistics**

**John E. Cremeans**  
**Office of Business Analysis**

**Zahava D. Doering**  
**Smithsonian Institution**

**Daniel E. Garnick**  
**Bureau of Economic Analysis**

**C. Terry Ireland**  
**National Security Agency**

**Charles D. Jones**  
**Bureau of the Census**

**Daniel Kasprzyk**  
**Bureau of the Census**

**David A. Pierce**  
**Federal Reserve Board**

**Thomas J. Plewes**  
**Bureau of Labor Statistics**

**Wesley L. Schaible**  
**Bureau of Labor Statistics**

**Fritz J. Scheuren**  
**Internal Revenue Service**

**Monroe G. Sirken**  
**National Center for Health**  
**Statistics**

**Robert D. Tortora**  
**National Agricultural**  
**Statistics Service**

## PREFACE

The Federal Committee on Statistical Methodology was organized by OMB in 1975 to investigate methodological issues in Federal statistics. Members of the committee, selected by OMB on the basis of their individual expertise and interest in statistical methods, serve in their personal capacity rather than as agency representatives. The committee conducts its work through subcommittees that are organized to study particular issues and that are open to any Federal employee who wishes to participate in the studies. Working papers are prepared by the subcommittee members and reflect only their individual and collective ideas.

The Subcommittee on Measurement of Quality in Establishment Surveys was formed to document, profile, and discuss the topic of quality in Federal surveys of establishments. In preparing this report, the Subcommittee walked in uncharted territory. Unlike the field of household surveys where there is a rich variety and depth of study in design and practice, the literature specifically pertaining to surveys of establishments is limited. The Subcommittee also found that the lack of a literature was reflected in a lack of standard practice amongst and within the agencies. It is hoped that this report will begin the process of narrowing the variations in design and practice as agencies are able to benefit from a profiling of establishment surveys.

Consequently, the Subcommittee report is presented in a format and style that aims to increase awareness on the part of sponsors and subject matter specialists of the major sources of error (sampling and nonsampling) associated with establishment surveys and to provide a basis for comparing agency survey procedures and practices with those of other agencies. When possible, observations are made in this report that would serve as a guide to planning and developing surveys with an appreciation for sources of error and a commitment to eliminating those sources to achieve quality in establishment surveys.

This report may also be of interest to a wider audience of those who collect information from establishments. To this end, the Subcommittee intends to organize seminars and meetings to discuss the topics with both Federal agency personnel and others in the broad statistical community.

The Subcommittee on Measurement of Quality in Establishment Surveys was chaired by Thomas J. Plewes of the Bureau of Labor Statistics, Department of Labor.

MEMBERS OF THE SUBCOMMITTEE ON  
MEASUREMENT OF QUALITY IN ESTABLISHMENT SURVEYS

Thomas J. Plewes\* (Chair)  
Bureau of Labor Statistics (Labor)

Kennon R. Copeland  
Bureau of Labor Statistics (Labor)

Carol Corby  
Bureau of the Census (Commerce)

Ronald S. Pecso  
National Agricultural Statistics Service (Agriculture)

Stanley R. Freedman  
Energy Information Administration (Energy)

Maria E. Gonzalez\* (ex officio)  
Office of Information and Regulatory Affairs (OMB)

Carl A. Konschnik  
Bureau of the Census (Commerce)

Samuel M. Slowinski  
Federal Reserve Board

Alan R. Tupek  
Bureau of Labor Statistics (Labor)

Preston J. Waite  
Bureau of the Census (Commerce)

George S. Werking  
Bureau of Labor Statistics (Labor)

\*Member, Federal Committee on Statistical Methodology

## ACKNOWLEDGEMENTS

This report represents an intensive effort on the part of dedicated Subcommittee members and outside reviewers over a two-year developmental period. It is truly a collective effort on the part of the members of the Subcommittee on Measurement of Quality in Establishment Surveys. The personal commitment of the members to the collective task was evident in the fact that several members continued their contribution despite changes in assignment that moved them from the positions in which they were employed at the time of the formation of the Subcommittee.

All members of the Subcommittee reviewed and approved the entire final report, but individual members had primary responsibility for the several chapters. At the time the sections were prepared in draft, the Subcommittee benefitted from an outside review of each section by members of the Federal Committee on Statistical Methodology who provided comments and suggestions that were invaluable in improving the final product. The names of the authors and expert reviewers of the several chapters of this report appear below.

<u>Chapter</u>	<u>Author</u>	<u>Reviewer</u>
I	Thomas J. Plewes	Maria E. Gonzalez
II	Kennon R. Copeland	Maria E. Gonzalez
III.A-C	Alan R. Tupek	Wesley L. Schaible
III.D-E	Preston J. Waite	Kirk M. Wolter
IV.A	Ronald S. Fecso	Robert D. Tortora
IV.B	Stanley R. Freedman	Yvonne M. Bishop
IV.C	Carl A. Konschnik	Daniel Kasprzyk
IV.D	Ronald S. Fecso	Robert D. Tortora
IV.E	Samuel M. Slowinski	David Pierce
IV.F	Carol Corby	Nash Monsour

Many persons deserve recognition when a Subcommittee completes its work. The list is extensive in the case of this report, though special recognition of the contribution of Maria E. Gonzalez during the gestation and production phases of the project cannot be overlooked. Her dedication to delivery of a quality product, on time, inspired the Subcommittee.

George S. Werking, Bureau of Labor Statistics, provided guidance to the Subcommittee in developing the organization of this report and the profile of survey practices.

Over its period of development, the report was twice presented to and commented upon by the Federal Committee on Statistical Methodology. Special appreciation is extended to Robert D. Tortora, National Agricultural Statistics Service and Wesley L. Schaible, Bureau of Labor Statistics, for their lead comments during these review sessions. Much of what finally appears in the report is in direct response to their suggestions and evidences their assistance.

The Subcommittee also expresses its appreciation to the many survey managers and designers across the agencies for their cooperation with the Subcommittee during the data-gathering operation. Without exception, those responsible for the Federal government's surveys of establishments take their work very seriously, are dedicated to providing quality data, are committed to improving their practices, and are intent on protecting the confidentiality of the data entrusted to their care while openly discussing their survey procedures.

A special word of appreciation is extended to Kennon R. Copeland of the Bureau of Labor Statistics who served as Secretariat for the Subcommittee and who personally conducted the data collection for the Federal agencies who were not represented directly by Subcommittee membership. Editing and typing services were ably provided by Editorial Experts, Inc., under contract with the Bureau of Labor Statistics for this purpose.

## TABLE OF CONTENTS

	Page
<b>Chapter I. EXECUTIVE SUMMARY</b>	1
A. Introduction	1
B. Survey Quality	1
C. Sample Design and Estimation	3
D. Survey Methods and Operations	3
E. Next Steps	4
 <b>Chapter II. BACKGROUND</b>	 7
A. Scope, Audience, and Objectives	7
B. Survey Quality and Subcommittee Approach to Report	8
C. Summary Profile of Survey Practices	8
D. Organization of Report	9
 <b>Chapter III. SAMPLE DESIGN AND ESTIMATION</b>	 11
A. Introduction	11
1. Basic Concepts	11
2. Reporting Unit: Establishment, Company, or Enterprise	11
3. Census versus Sample	11
4. Probability versus Nonprobability	12
B. Establishment Universe Populations and Frames	13
1. Background	13
2. Establishment Population Distribution	13
3. Sample Frame Approaches	14
4. Common Characteristics of Establishment List Frames	14
5. Maintaining a Frame	15
C. Sample Design	16
1. Background	16
2. Common Characteristics of Sample Designs	16
3. Sample Redesigns	18
4. Summary Profile	19
D. Estimation	23
1. Background	23
2. Commonly Used Estimators	23
E. Sampling Error Estimation	28
1. Background	28
2. Common Approaches to Variance Estimation	28
3. Factors Affecting the Use of Variances in Establishment Surveys	30
4. Summary Profile	31

	Page
Chapter IV. SURVEY METHODS AND OPERATIONS	33
A. Introduction	33
1. Basic Concepts	33
2. Error Measurement	33
B. Specification Error	34
1. Definition of Specification Error	34
2. Sources of Specification Error	34
3. Control of Specification Error	38
4. Measurement of Specification Error	39
5. Summary Profile	40
C. Coverage Error	44
1. Definition of Coverage Error	44
2. Sources of Coverage Error	44
3. Control of Coverage Error	48
4. Measurement of Coverage Error	50
5. Summary Profile	51
D. Response Error	57
1. Definition of Response Error	57
2. Sources of Response Error	58
3. Control of Response Error	59
4. Measurement of Response Error	61
5. Summary Profile	63
E. Nonresponse Error	68
1. Definition of Nonresponse Error	68
2. Sources of Nonresponse Error	68
3. Control of Nonresponse Error	70
4. Measurement of Nonresponse Error	72
5. Summary Profile	74
F. Processing Error	79
1. Definition of Processing Error	79
2. Sources of Processing Error	79
3. Control of Processing Error	81
4. Measurement of Processing Error	82
5. Summary Profile	82
REFERENCES	86
APPENDIX 1. Goals, Scope, and Uses	89
APPENDIX 2. Survey Profile Questionnaire	90
APPENDIX 3. Profile of Survey Practices: Federal Establishment Surveys Covered	101

## LIST OF FIGURES

Figure 1.	Survey Program Requirements	20
Figure 2.	Sample Design	21
Figure 3.	Estimation	32
Figure 4.	Specification Error Control Procedures	42
Figure 5.	Specification Error Measurement Techniques	43
Figure 6.	Coverage Error Control Procedures	53
Figure 7.	Coverage Error Measurement Techniques	55
Figure 8.	Response Error Control Procedures	64
Figure 9.	Response Error Measurement Techniques	66
Figure 10.	Nonresponse Error Control Procedures	76
Figure 11.	Nonresponse Error Measurement Techniques	78
Figure 12.	Processing Error Control Procedures	84
Figure 13.	Processing Error Measurement Techniques	85

## CHAPTER I. EXECUTIVE SUMMARY

### A. INTRODUCTION

Data collected in surveys and censuses of establishments comprise an integral and important part of the nation's information base for policymaking and analysis. Key information on employment and wages, sales, prices, agriculture and energy production, money supply, and many other aspects of the working of the economic and social order are collected from businesses, compiled and published by a large number of Federal government agencies.

The collection of data from establishments is not new. Some of the establishment-based data series have been continuous since the early part of this century, and many predate household surveys. Nonetheless, in contrast with household surveys, for which a rich literature has emerged over the past 5 decades, very little in the way of theoretical or evaluative work on survey quality has been published for establishment surveys.

The comparative shortage of literature and the government's approach to establishment surveys have resulted in a situation unique to establishment surveys. Today, there are few commonly accepted approaches to the design, collection, estimation, analysis, and publication of establishment surveys. Establishment surveys abound in rich variety, with little standardization of design, practice, and procedures.

This is not to say that Federal agencies do not work hard to ensure that the surveys they conduct are carried out in the most professional and efficient manner that is possible given the resources available. The members of this subcommittee, the agencies they represented, and the representatives of agencies interviewed for this study were serious in their efforts to ensure the quality of their products. They do so not only because they want to, but because they are obliged to do so by the Office of Management and Budget's clearance process. However, both the agency personnel that have responsibility for the establishment surveys and the OMB staff that reviews the requests for new and renewal surveys operate without benefit of key design information available from a profile describing the quality of surveys. The collectors and reviewers, and more importantly the users of establishment data, would be greatly assisted if there were a better understanding of the sources of error in the surveys and censuses, and a sharing of information on methods for dealing with or overcoming those error sources to achieve higher quality data.

### B. SURVEY QUALITY

This report discusses, in very general terms, the potential sources of error that may affect the quality of counts and estimates derived from surveys and censuses of establishments. By classifying these sources of error, the report focuses on practices that are used to improve and measure the quality of establishment data. To this extent, the approach of the Subcommittee on Measurement of Quality in Establishment Surveys

was rather straightforward and fairly conventional. For example, only the more traditional aspects of quality are considered--those that refer to the accuracy of the survey estimate or its closeness to a "true" value. Other aspects of quality such as relevance and timeliness, which the current literature considers to be critical components of a total quality approach from the vantage point of the user, are not given equal emphasis.

The report retains the usual distinction between sampling error and non-sampling error as the central dichotomization. Sampling error is discussed in terms of sample design, estimation, and variance estimation. The survey methods and operations determine nonsampling errors which are partitioned into five areas--specification error, coverage error, response error, nonresponse error, and processing error. Error is discussed in terms of sources, control, and measurement.

As part of the discussion of survey quality, contrasts between establishment and household surveys are mentioned. There are very real and, in some instances, major differences in sources of error. Household surveys do not have to worry about complex corporation structures and affiliations, free trade zones, Government versus private ownership, onshore versus offshore activities, definitional differences such as gas bought and sold versus transported, etc. All of these issues serve to complicate the control and measurement of error in establishment surveys.

The core of this study is a profile of the Federal government's current establishment survey environment. In an attempt to quantify the information presented in the report, the Subcommittee collected data on design, estimation, control, and measurement practices for 55 surveys from 9 Federal agencies. The surveys were selected to include a large number of the known major ongoing establishment surveys conducted by the Federal government and thus provide a comprehensive snapshot of the current establishment survey environment.

Key points from the discussion of establishment survey error sources, control, and measurement are summarized below. Three major points are worth stating as a premise to the summary:

- In general, establishment surveys have procedures in place designed to control major known sources of survey error;
- Error measurements are not extensively derived;
- Error measurements are seldom published when they have been estimated.

While the relative differences in the extent of use of control and measurement can be understood in terms of resource priorities, there does not appear to be a clear reason why error information is not published when available. The limitations in the availability of published error information made it quite difficult for the Subcommittee to collect this information. Hopefully, now that collection has been completed, this report will become more valuable as a reference document.

### C. SAMPLE DESIGN AND ESTIMATION

Establishments are different from households. The distributions of their populations are very skewed, with a few large firms commonly dominating totals for most characteristics of interest. These distributions affect the frame development and maintenance, sample design, and estimation practices of establishment surveys. Given the importance of large units, extensive resources are devoted to improving frame coverage and content for large units. One-stage, highly stratified designs, with certainty selection of large establishments are used in the vast majority of establishment surveys profiled.

About four-fifths of the surveys profiled were designed and implemented as probability surveys. Roughly one-fifth of the surveys profiled were described as having designs or implementations which do not result in a probability design. These surveys included those for which substitution is allowed for nonresponse, a segment of the target population has no chance of selection, units are selected judgmentally, and other practices are followed that are at variance with probability design practice. Cost versus quality tradeoffs were often cited as reasons for deviations from common probability design/implementation.

Estimators which do not reflect probability of selection are also commonly used in establishment surveys. Those estimators may generally be described as model-based, although the model often is implicit, rather than explicitly stated. Estimates for small firms are frequently derived using administrative data or data from larger firms, because cutoff sampling is used in about one-fourth of the surveys.

One-fourth of the sample surveys profiled in the data collection by the Subcommittee did not compute variances, and another one-fifth did not publish estimates of sampling error in survey publications. This lack of generation and publication of sampling error information was not seen to be a function of agency practice, since it was not confined to one or two agencies, but rather it appeared to be somewhat correlated with the use of nonprobability-based estimation procedures.

### D. SURVEY METHODS AND OPERATIONS

Establishment surveys typically seek hard data for which records are available. This is a central characteristic which both simplifies the collection and complicates the interpretation of the data. The collection is simplified because there are hard data on record from which the data of interest are extracted, rather than relying on the memory, opinions, or interpretations of the respondents as is often the case for household surveys. The survey methods and operations used determine nonsampling errors affecting the quality of the resulting data.

However, in establishing the concepts and definitions to be used in the surveys, special care must be taken to consider carefully the establishments' recordkeeping systems, definitions, and data availability to avoid introducing specification error into the data. Typically, agencies do this through a requirements review or a respondent or trade association

consultation. How well the agencies perform this function is difficult to measure. There is currently no single specification error measurement practice used by a large majority of the surveys profiled. Slightly more than half of the surveys regularly compared survey results to independent estimates to gain a better understanding of specification error.

Establishment surveys commonly use list frames, and thus are subject to the inherent problems associated with list frames--duplication, over-coverage of out-of-scope and out-of-business units, undercoverage of business births, and misclassification of units. In apparent recognition of these potential sources of error, well over half of the surveys profiled regularly used procedures designed to control these problems, such as updating for structural changes, updating/sampling for births, and internal consistency checks for duplicates. On the coverage error measurement side, little is commonly done except to provide such indirect measures as out-of-business and out-of-scope rates. No direct measurement technique was reported as regularly used by more than half of the surveys.

The fact that data are acquired from records also affects the sources of response error in establishment surveys while enabling subject-matter analysts to identify possible reporting error at the microdata level. As a result, common control procedures for response error include not only those typically in place for household surveys, such as editing for reasonableness, questionnaire pretest, and detailed training/guidelines for interviewers, but also include analyst review of data, and record-keeping practices studies. Outside of the calculation of edit failure rates, little response error measurement is done across surveys.

The control of nonresponse in establishment surveys generally relies upon conventional practices, including unit and item nonresponse followup, and advance notification. However, the skewed nature of the population has led to other widely used control techniques weighted toward large units which are unique to establishment surveys. These techniques include intensive followup of critical units, central office consolidation of all responses from the same establishment, other special reporting arrangements, and provision of survey publications to respondents. Several indirect measures of nonresponse error, such as unit and item response rates and refusal rates, are commonly generated. Because of the population distribution, weighted response rates are also commonly derived. Very little is done on direct measurement of nonresponse error.

Control procedures for processing error do not differ from those in use for household surveys. The identified control procedures were all used by over half of the surveys profiled. The most common measurements produced were edit failure rates, which, as noted earlier, are generated from concern about response error as well as about processing error.

#### E. NEXT STEPS

No specific recommendations are made in this report. The Subcommittee trusts that the discussion and profiling of error sources as applied to

establishment surveys will give impetus to consideration of survey practices on the kind of case-by-case basis that is necessary given the vast differences in the establishment survey operations. Nonetheless, the tenor of the findings can be depicted as recommending more work to improve and document the quality of surveys. The profile portrays a number of key Federal government surveys with deficiencies in the measurement and documentation of sampling and nonsampling errors, and points to a need to focus additional attention, and resources, on the general improvement and documentation of survey practices.

The profile has also reminded us of the limitations of our understanding of errors, their sources, and the means of reducing or accounting for them. More importantly, little is known of the interaction of the errors. To the extent that this profile engenders interest in continuing this common exploration, it will have more than proved its usefulness.

On the positive side, the Subcommittee believes that the framework that has been adopted here--an amalgam of theory and practice--provides a useful tool for a systematic approach to understanding and evaluating quality in establishment surveys. It constitutes a step in the process of quantifying and improving the quality of the important surveys of establishments conducted by the Federal government.

In addition, the Subcommittee plans to organize seminars to discuss this report with Federal agencies. These seminars should serve to promote a greater interest among Federal agencies in analyzing and improving the quality of the establishment surveys they sponsor.

## CHAPTER II. BACKGROUND

### A. SCOPE, AUDIENCE, AND OBJECTIVES

The Federal government sponsors, conducts, and publishes data from a number of surveys of establishments in the United States. These surveys provide a wealth of information about the economic well-being of the country for government policymakers and the business community. Although there is some overlap of survey design issues between establishment and household surveys, there exist a number of important differences between the two. Much has been written about survey design issues associated with household surveys. The extent of literature available for establishment surveys, however, is limited.

The Subcommittee on Measurement of Quality in Establishment Surveys was established by the Federal Committee on Statistical Methodology in November 1985 to document, profile, and discuss the topic of quality in Federal surveys of establishments. The Subcommittee established the following goals for its report:

- Document current understanding of the meaning of quality in establishment surveys;
- Discuss establishment surveys in terms of sampling and nonsampling error;
- Identify approaches and practices to be considered by users and designers of establishment surveys;
- Profile current practices in the areas of controlling and measuring survey quality.

Although the objectives of the Subcommittee were quite broad, the scope of its work was narrowed early to a manageable slice of a very large Federal undertaking. Thus, while the Subcommittee sought to be encompassing in focusing on all Federal agencies that conduct or sponsor surveys of establishments, the range of experience brought into the discussion was necessarily limited to the membership of the Subcommittee. Information concerning practices in other agencies were incorporated into the report through the profile of current practices. The scope of surveys profiled was restricted to ongoing surveys of private sector establishments. Establishment was interpreted in the broadest sense to include corporations, partnerships, and sole proprietorships engaged in agriculture, mining, construction, manufacturing, trade, and/or services. One-time surveys, special studies, and surveys covering only government establishments were excluded for both practical reasons and priority of interest.

This report is intended to provide reference and guidance for survey practitioners--statisticians, survey managers, analysts, and agency policymakers--across the Federal government in planning and refining establishment surveys. The report does not attempt to define standards nor to evaluate the current practices used in particular surveys.

A more detailed list of the goals, scopes, and uses of the report that were developed by the Subcommittee to serve as a guideline for the development of the report is provided in Appendix 1. This report represents the results of the Subcommittee's effort toward achieving those goals initially set forth.

#### **B. SURVEY QUALITY AND SUBCOMMITTEE APPROACH TO REPORT**

The Subcommittee translated the notion of quality into the topic of errors associated with survey estimates. A survey design consists of a sampling plan (sample design), estimation procedures, and survey methods and operations (including development of a frame, design of a questionnaire, data collection procedures, and processing operations). Each of these components may contribute to the error in the resulting survey estimates. Thus even a census, which requires no sampling plan nor estimation procedures, is subject to errors of measurement resulting from the survey procedures used.

Survey estimates are subject to both variable error and bias. Variable error reflects random error resulting from the survey design and conduct, while bias reflects systematic error. More detailed discussion of the models available to represent survey errors may be found in most sample theory textbooks, such as Cochran (1977), Kish (1967), and Hansen, Hurwitz and Madow (1953).

Errors resulting from the sample design and estimation are referred to here collectively as sampling error, while errors resulting from the survey methods and operations are referred to as nonsampling error. These two components defined the structure for discussion of survey error. Discussion of establishment universe populations was included in the first part to provide the context for sample design and estimation. Nonsampling error was partitioned into five areas by the Subcommittee-- specification error, coverage error, response error, nonresponse error, and processing error.

A Subcommittee member was assigned to write a section for each of the areas identified. In a series of meetings, Subcommittee members exchanged ideas and individual and agency experiences. The structure of those meetings was to first discuss ways in which errors can arise in the course of a survey. Following that, methods used to control those sources of error were discussed. Finally, measurements obtained to provide information about errors were discussed. These meetings resulted in a framework for the paper, and an identification of the information to be collected for the profile of quality in establishment surveys.

#### **C. SUMMARY PROFILE OF SURVEY PRACTICES**

Information on survey design practices was collected to complement the discussion contained in the report. A questionnaire was developed to allow Subcommittee members to collect information on sample design, estimation, and control and measurement techniques. Appendix 2 contains the questions and items collected, along with explanations provided for the list of control procedures.

Subcommittee members identified surveys within their respective agencies to be profiled. In addition, four agencies not represented on the Subcommittee (National Center for Education Statistics, Bureau of Economic Analysis, Bureau of Mines, National Center for Health Statistics) were contacted and surveys identified for collection of data. The Subcommittee collected information on the survey design practices of 55 Federal establishment surveys from nine agencies (see Appendix 3).

Collection of data for the represented surveys was carried out by the Subcommittee members in consultation with responsible staff at their respective agencies. Data for the nonrepresented agencies was collected by one of the Subcommittee members through interviews with appropriate statisticians and survey managers at the agencies.

The data obtained are summarized in the figures appearing in the report and are discussed in the summary profile sections. Unless stated otherwise, the base for the percentages is the 55 surveys covered by the survey profile questionnaire. The data were collected to provide a summary profile of the current Federal establishment survey environment, not to profile or compare individual survey practices. The data have not undergone the formal agency review and clearance which would be required to publish or release information about specific surveys.

The figures for the five nonsampling error sections present the data similarly. First, the control procedures are presented in decreasing order of frequency of use. Frequency of use is classified by usage on a regular basis (solid portion of bar) or an irregular basis (cross-hatched portion of bar). Some procedures are not applicable (N/A) for certain surveys (e.g., reinterview sample of interviewer's work for mail only surveys). The frequency, if any, of non-applicable procedures are indicated by the white portion of the bar. The space between the top of the bar and 100% represents non-usage of the procedure.

Second, the measurement techniques are presented (indirect measures followed by the direct measures) in decreasing order of frequency of use. The bars for each technique have two sides. The left side represents the frequency of use--regular basis (solid) or irregular basis (cross-hatched)--and the right side represents the application of the measures obtained--internal use only (solid) or published (cross-hatched). As for the control procedures, not applicable is indicated by the white portion of the bar, and non-usage of the measurement technique is the space between the top of the bar and 100%.

#### D. ORGANIZATION OF REPORT

The remainder of the report contains two chapters. Chapter III contains approaches to and issues associated with sample design and estimation. Chapter IV contains discussion of sources of error, control techniques, and measurement techniques for the five components of nonsampling error as defined by the Subcommittee. Following discussion of each topic within the chapters, summary profile data obtained from the survey of Federal establishment surveys are presented.

## CHAPTER III. SAMPLE DESIGN AND ESTIMATION

### A. INTRODUCTION

#### 1. BASIC CONCEPTS

This chapter focuses on frame, sample design and estimation approaches for establishment surveys, and the resultant sampling error. A frame is a list of units which makes up the population (Cochran, 1977). The sample design, as used in this report, refers to that part of the survey design which includes the organization of the frame and method of choosing the sample (sampling plan). Estimation refers to the methodology used to generate estimates for the population based on the sample data. Sampling error can be defined as that part of the difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a sample of values is observed (Kendall and Buckland, 1960). In general, an estimate of the sampling error can be derived from the particular sample selected for the survey.

#### 2. REPORTING UNIT: ESTABLISHMENT, COMPANY, OR ENTERPRISE

A reporting unit designates the unit for which data are to be collected. Survey data are usually collected at the establishment level. An establishment is not necessarily identical with an enterprise or company, which may consist of one or more establishments. Also, it is to be distinguished from subunits, departments, or divisions (Office of Management and Budget, 1987). An establishment is usually defined as an economic unit, generally at a single physical location, where business is conducted or services or industrial operations are performed. Survey data are occasionally collected at the enterprise or company level such as for surveys of U.S. enterprises owning foreign subsidiaries (Bureau of Economic Analysis), or for surveys of corporations' financial reports (Bureau of the Census).

#### 3. CENSUS VERSUS SAMPLE

A complete enumeration or census of all units on the frame is not unusual (approximately one-sixth of the surveys profiled) for establishment surveys. Many surveys are conducted for a particular industry or area of the country where there are so few units that a census is both feasible and efficient. While a census is not subject to sampling error, both censuses and surveys are subject to nonsampling errors. Nonsampling error can be attributed to a variety of sources resulting from the survey design: inability to obtain information about all cases in the sample; definitional difficulties; differences in the interpretation of questions; inability or unwillingness to provide correct information on the part of respondents; mistakes in recording or coding the data obtained; and other errors of collection, response, processing, coverage, and estimation for missing data (U.S. Bureau of the Census, 1974). Sources, control and measurement of nonsampling error are discussed in Chapter IV.

#### **4. PROBABILITY VERSUS NONPROBABILITY**

A number of Federal establishment surveys were not classified as probability sample designs (approximately one-fifth of the surveys profiled), based on the definition developed by the Subcommittee. Survey managers were asked to classify their survey as nonprobability if one or more of the following conditions existed: substitution is allowed for nonrespondents; some large set of units in the target population have no chance of selection; units are selected judgmentally; no adequate frame exists; sample too hard to control; other--specify. Some of these conditions indicate a nonprobability design, while others indicate lack of control in implementing the design. The nonprobability surveys were found in almost all statistical agencies. In most situations, survey managers cite cost/quality tradeoffs as reasons for nonprobability sample design. Also, nonprobability samples may have been selected many years ago and the sample design has not been updated.

## **B. ESTABLISHMENT UNIVERSE POPULATIONS AND FRAMES**

### **1. BACKGROUND**

Establishment populations differ from household populations in several ways. These dissimilarities result in frame development, sample design, and estimation approaches which are in some areas markedly different from approaches for household surveys. Among the major distinctions between establishment and household populations and frames are: (1) establishments come from skewed populations wherein units do not contribute equally (or nearly equally) to characteristic totals, as is the case for households; and (2) accuracy of frame information about individual population units is crucial to sample design and estimation for establishment surveys, while for household surveys the accuracy of frame characteristics concerning individual units is not as critical to the sample design.

### **2. ESTABLISHMENT POPULATION DISTRIBUTION**

Establishment surveys are characterized by the skewed nature of the establishment population (see, for example, Table 1). A few large firms commonly dominate the estimates for most of the characteristics of interest. This is especially true for characteristics tabulated within an industry. Small firms may be numerous, but often have little impact on survey estimates of level although they may be more critical to estimates of change over time or for measuring characteristics related to new businesses. This distribution has a major impact on both the frame development and maintenance and on the sample designs used for establishment surveys.

Table 1

---

Distribution of Establishments on the Bureau of Labor  
Statistics List Frame by Number of Employees  
(First Quarter, 1987)

---

SIZE CLASS (No. of employees)	% OF ALL UNITS	% OF ALL EMPLOYEES
ALL	100.0	100.0
0 - 4	58.3	6.5
5 - 9	18.1	7.8
10 - 19	11.1	9.8
20 - 49	7.5	14.9
50 - 99	2.7	12.4
100 - 249	1.6	15.5
250 - 499	0.4	9.7
500 - 999	0.2	8.0
1000+	0.1	15.4

---

SOURCE: U.S. Bureau of Labor Statistics

---

### **3. SAMPLE FRAME APPROACHES**

#### **List Frames**

List frames are widely used in establishment surveys conducted by the Federal government. The use of list frames for establishment surveys arose from the availability of administrative records on businesses compiled mainly for tax purposes. Theoretically, all businesses must pay (or justify not paying) Federal, State, and local income taxes (where applicable), social security tax, unemployment insurance tax, and other taxes. Filing requirements of State and Federal Government agencies provide the conceptual basis for frame coverage of business establishments. In addition, regulatory reporting requirements provide lists of establishments in certain industries, such as oil refineries. However, because these administrative record files are not normally developed for statistical purposes, they often need refinement before being used as sampling frames for surveys of businesses. Thus addresses used for administrative purposes may not be adequate for survey purposes. For example, an address in the administrative files could be for the accounting firm that handles tax reports for the company on the list frame. Extensive resources are spent on maintaining the list frames since a significant source of non-sampling error may be due to inadequacies in the frame. Resources for improving frame coverage and the accuracy of identification data are typically spent on improving the data for the larger firms since they have a much greater impact on most survey estimates. Procedures for improving the quality of list frames are discussed in Section IV.C.

#### **Area Frames**

While most establishment surveys use list frames, surveys conducted by the Department of Agriculture rely heavily on area sampling in combination with list frames. Retail Trade Surveys conducted by the Bureau of the Census use an area sampling frame to supplement their list frame. Area sampling frames have the advantage of complete coverage of even new businesses. However, the costs involved in changing the stratification for an area frame limit the frequency with which sample design modifications can be made to reflect changing population distributions. Area frames are therefore more efficient when used on stable populations, such as agriculture.

### **4. COMMON CHARACTERISTICS OF ESTABLISHMENT LIST FRAMES**

Establishment list frames typically are characterized by extensive establishment identification information; periodic updating of this information, and multiple sources for the information. Information usually includes the name and address of the establishment, industry and ownership codes, size data (employment, sales, enrollment, etc.), a unique identification number, a link to related establishments, and other data items specific to the surveys that the frame must service. The data on the frame are required for sample design, sample selection, identification of sample units, and estimation. The primary source of administrative records for a frame may have shortcomings which require the identification information to be supplemented using other sources of information. This

may include using identification information from the surveys themselves. Supplemental files, including the use of area frames, may also be required to overcome coverage problems in the primary source. Duplication of sampling units is also a problem associated with the use of list frames. Refinement of the frame includes efforts to unduplicate units prior to sampling.

## 5. MAINTAINING A FRAME

The individual establishment information on the frame is critical to the effectiveness of the sample design and estimation for the survey. Maintaining a frame over time is complicated by the dynamic nature of the establishment community. Changes in ownership, mergers, buyouts, and internal reorganizations make frame maintenance a real challenge. Matching and maintaining unit integrity over time provides the opportunity for consistent unit identification in the numerous periodic surveys conducted by the Federal Government.

New establishments must be added to the frame. However, it is often difficult to differentiate, using administrative records, new establishments from old establishments that have changed their name or corporate identity. It is also difficult to link businesses over time when there have been ownership or other changes. Each survey may have different requirements as to the handling of new establishments and changes in existing establishments. The timeliness of adding new establishments to the frame and reflecting them in the sample is also a problem. The lag time between formation of new establishments and selecting them into the sample may be anywhere from several months to several years. While new establishments may have little impact on estimates of level, in some instances they may dominate estimates of change.

The Bureau of the Census and the Bureau of Labor Statistics both have independent programs for maintaining frames for large and multiunit companies, since provisions for confidentiality prevent sharing between agencies. The Census Bureau conducts an annual Company Organization Survey to determine and maintain the structure of business enterprises. The Bureau of Labor Statistics through cooperating State Employment Security Agencies conducts a quarterly survey of identified multiunit companies to determine units that have been bought, sold, or merged. These surveys are necessitated because there are as many as 800,000 new non-agricultural employers each year, up to 5 percent of existing establishments may change industry classification, and the number of mergers is steadily increasing.

## **C. SAMPLE DESIGN**

### **1. BACKGROUND**

Establishment surveys differ from household surveys in the sample design approaches taken. Establishment surveys typically use single-stage designs, as opposed to the multistage designs typical for household surveys. The dominance by a small set of units on estimates of characteristics of interest leads to differential sampling by establishment size, with the use of certainty strata beyond that determined by the optimal allocation. The use of certainty strata is often to protect against the possibility of inefficiencies in the design parameters. Overlap of sample units across survey rounds is often optimized to improve estimates of change and reduce collection costs and nonresponse rates. These situations correspond to those found for household survey primary sampling units (PSUs), which typically have differential and certainty sampling as well as overlap of PSUs across survey rounds.

### **2. COMMON CHARACTERISTICS OF SAMPLE DESIGNS**

Establishment surveys have similarities in sample design approaches as well as frame approaches. The approaches are due to the distribution of the population and the amount of unit information available on the frame. A typical establishment survey sample design is a single-stage, highly stratified design. Stratification is by industry, size (employment, sales, etc.), and/or geographic location. The larger units are selected with certainty, and very small units may either be excluded from the target population or be given no chance of selection. Sampling within strata is either equal or probability proportional to size.

Administrative record data are often used as design variables for stratification and allocation. The administrative record data from the Internal Revenue Service, Social Security Administration, State Unemployment Insurance Agencies, and other sources may agree with survey definitions, but they are often not timely enough for survey schedules. The accuracy of data is undoubtedly a function of how critical the data values are to the administrative source collecting them. But even when administrative records are untimely or somewhat imprecise, they are often valuable as design characteristics. For example, the Census Bureau uses race and sex codes from administrative records on the owners of sole proprietorships and partnerships to aid in developing a very efficient sample design for the Survey of Minority Owned Businesses.

Establishment surveys are often stratified first by geography and industry since separate estimates are often produced by geographic region and by industry. Even when geographic and industry breakouts are not produced, differences in the design variables by geographic area or industry may justify this stratification. A size measure such as employment or sales is often the most critical stratification variable. Since characteristics to be estimated are often highly correlated with the size measure, the use of the distribution of the size measure for stratification and allocation provides a highly efficient sample design.

Most survey estimates are dominated by characteristics of a few large firms; hence almost all designs sample more heavily from larger firms than from smaller firms, with most designs having certainty selection of the largest firms. The largest establishments will likely be in a "take all" stratum when optimum stratification techniques are used. In practice, a certainty stratum is often employed even when the allocation may not dictate it because a certain amount of protection is needed from imprecise design variables. Also, a standard certainty size class stratum may be employed across industries and geographic areas, rather than allowing the allocation to be determined by the design variables.

The importance and dominance of large firms have given rise to some non-classical designs. The smallest establishments may not be given a chance of selection since they contribute only marginally to the total estimate, are often covered inadequately on the frame, have erroneous data, are costly to collect, and tend to be volatile. A number of establishment surveys employ a form of cutoff sampling where no units are selected below a specified size. Data for smaller firms are either imputed from administrative records or from large firm characteristics, or they are excluded from the target population altogether. Obviously surveys that purport to cover all establishments must adjust for units not given a chance for selection.

In the Occupational Employment Survey conducted by the Bureau of Labor Statistics, units with less than four employees are not usually selected in the sample. Instead, the assumption is made that the occupational distribution of these units is the same as units responding in the next larger size class (four to nine employees). Similarly, the M3--Manufacturers' Shipment Inventories and Orders Survey--conducted by the Census Bureau does not sample units having fewer than 100 employees. Imputation for these units is also based on responses from the larger units.

The allocation of the sample will usually vary considerably by size of establishment. Units slightly smaller than the certainty cutoff will be given a much higher chance of selection than the smallest units. It is also common for designs to include differential target errors for the various industry and geographic estimating cells. This may be due to tradeoffs in the design between aggregate and detailed level estimates as well as to cost considerations. Small or volatile industries would command a significant portion of the sample if all estimating cells had a common target error.

Conflicting design objectives are common for establishment surveys, as is true for many household surveys. Tradeoffs exist between the need for detailed publication cells, limited or inefficient population design parameter data for detailed cells, and the survey cost related to increasing sample size. The sample design needed for detailed publication cells often increases the size of the sample significantly, with little gain in reliability in the aggregate cells. As an example, surveys conducted by the Bureau of Labor Statistics in cooperation with State Employment Security Agencies are intended to produce national as well as State estimates, and may be designed to produce sub-State estimates as well.

Establishment surveys are conducted monthly, quarterly, annually, and sometimes less frequently. Annual surveys often select independent samples from one year to the next. However, a number of surveys conducted by the Federal government use the same panel of units over time. Although estimates of level are the primary objectives of most surveys, estimates of change are also important. The use of a panel sample over time can improve the reliability of estimates of change for a given sample size. Panel units do not have to be reinitiated into the sample, lowering costs and increasing response rates. Household surveys view length of time in sample as a possible detriment to quality, due to the decreased response rates and the potential for conditioning effects on respondents. Given the hard data sources expected for establishment surveys (see IV.D), once a unit is used to reporting data under the definitions required for a survey, extended length of time in sample may not be a detriment to data quality.

Periodic establishment surveys often have special requirements which impact sample design and selection. These may include the need for large sample overlap from one survey round to the next or the need to minimize the sample overlap between survey rounds. Requirements such as these are intended to reduce the workload for the data collection staff, improve response rates, or reduce the burden on individual small establishments. To accommodate these and other requirements, rotating panel designs are used, or modifications are made to the independent sample selection of units from one survey round to the next. Even when independent samples are drawn, a large overlap in sample members is not uncommon due to the certainty size cutoff and the selection of a dense sample of larger firms.

### 3. SAMPLE REDESIGNS

Redesigning the survey periodically is an integral part of the survey process. Design objectives, population characteristics, survey resources, and features of the frame change over time. Requirements for survey estimates may change as funding changes or as the demand for estimates at various levels changes (discussed in IV.B). The growth and decline of various industries can also affect the criteria used for the sample design. Moreover, the availability of frames and the information on these frames may necessitate a complete redesign of the survey. Updates to the current design, including partial reselection of samples and revision of original probabilities of selection, may be adequate for a period of time, but eventually a redesign is essential.

A number of issues must be considered during the redesigning of the survey, such as continuity of the data series, the ability to analyze and the availability of data for determining the sample design, and the cost of the redesign relative to the ongoing survey. Maintaining the continuity of the data series requires a great deal of attention since the usefulness of the data may be due to its longitudinal aspects as much as it is to current measurement. Parallel processing under two designs is not uncommon, and helps ease the transition between designs.

Redesigns are often built into the survey process based on the recurrence of new frames or censuses. The economic censuses conducted by the Census Bureau every 5 years provide an opportunity for redesign of their

periodic surveys. The redesign of surveys may be conducted on an as-needed basis, such as when the current design is deemed inefficient or when more flexibility in the design is desired.

#### 4. SUMMARY PROFILE

(See Figures 1, 2a, and 2b.) Perhaps the most striking result obtained from the information on program requirements and sample design for the in-scope surveys is the extent of nonprobability sample designs, approximately one-fifth of the surveys (one-fourth of the sample surveys). Some surveys do plan probability sample designs, but in the course of sample selection, data collection, estimation, etc., control of the sample in terms of a probability design is lost. Others are designed as nonprobability by excluding a large portion of the target population, or using judgmental selection of units.

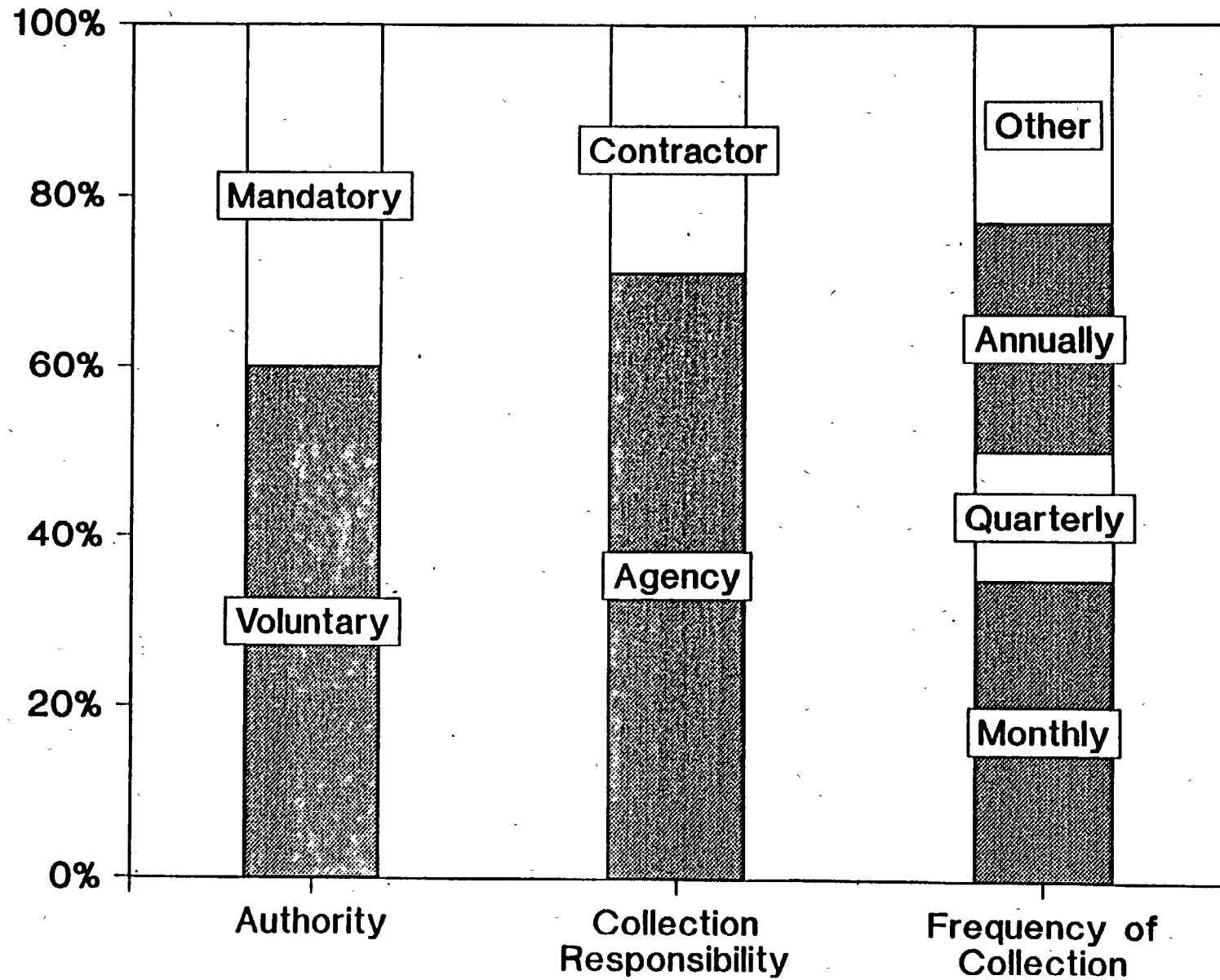
Approximately half of the nonprobability surveys were classified as such due to the design rather than due to implementation difficulties. Several surveys spanning most of the major statistical agencies used cutoff sampling, or judgmental sample selection. The other half of the nonprobability surveys were designed on a probability basis, but were not controlled in a manner the Subcommittee defined as probability (substitution for nonresponse, probability of selection not used, other control problems).

Approximately four-fifths of the sample surveys use certainty levels (e.g., all units above a designated size are included in the sample with certainty). Approximately 30 percent have sample cutoffs (e.g., all units below a designated size have no chance of selection). Some of the surveys do not include units below the sample cutoff in the target population while other surveys, as mentioned above, do include units below the sample cutoff in the target population.

Over four-fifths of the sample surveys have only one stage of selection. This is in contrast to household surveys which typically use multi-stage sample designs.

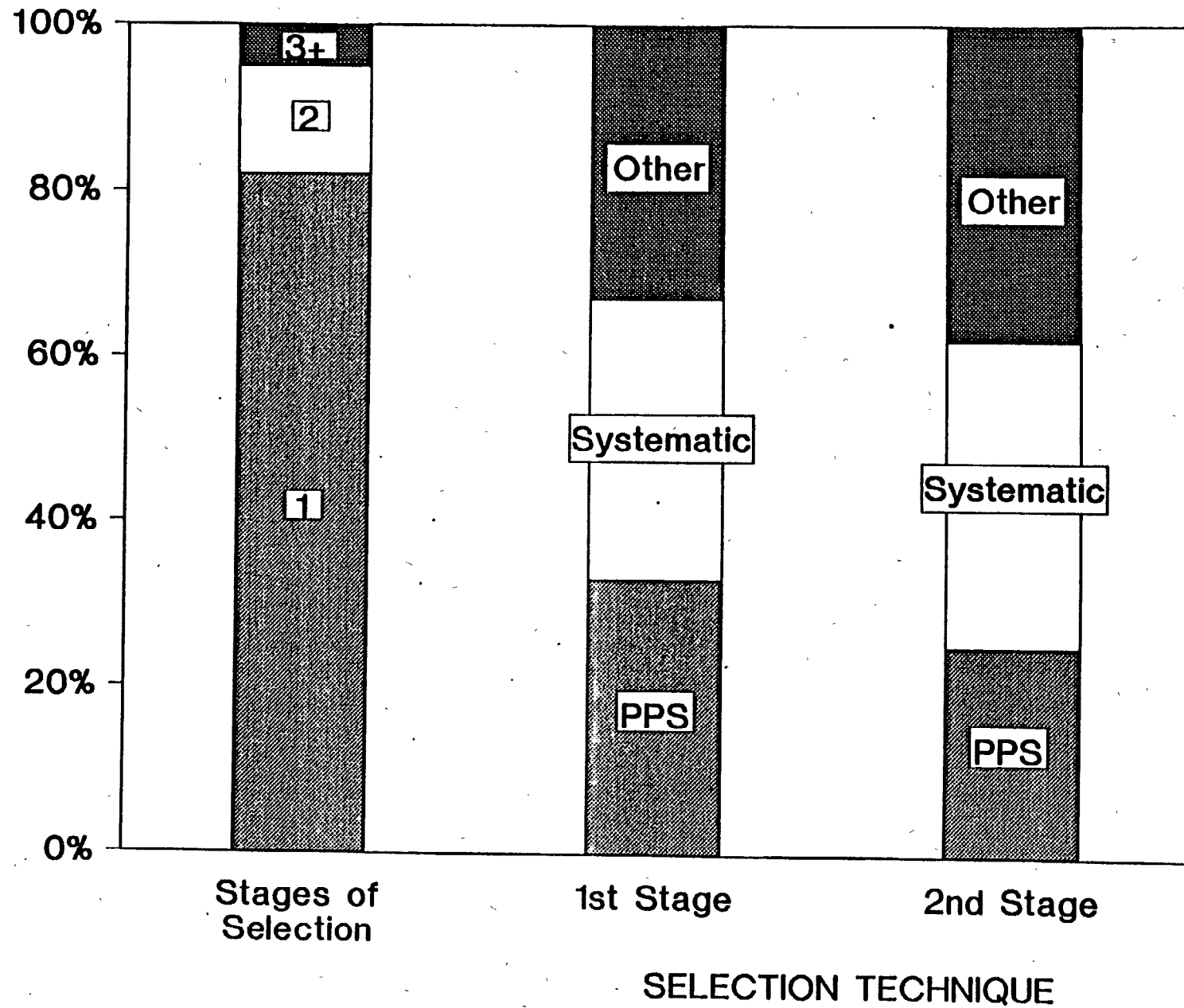
# SURVEY PROGRAM REQUIREMENTS

Figure 1



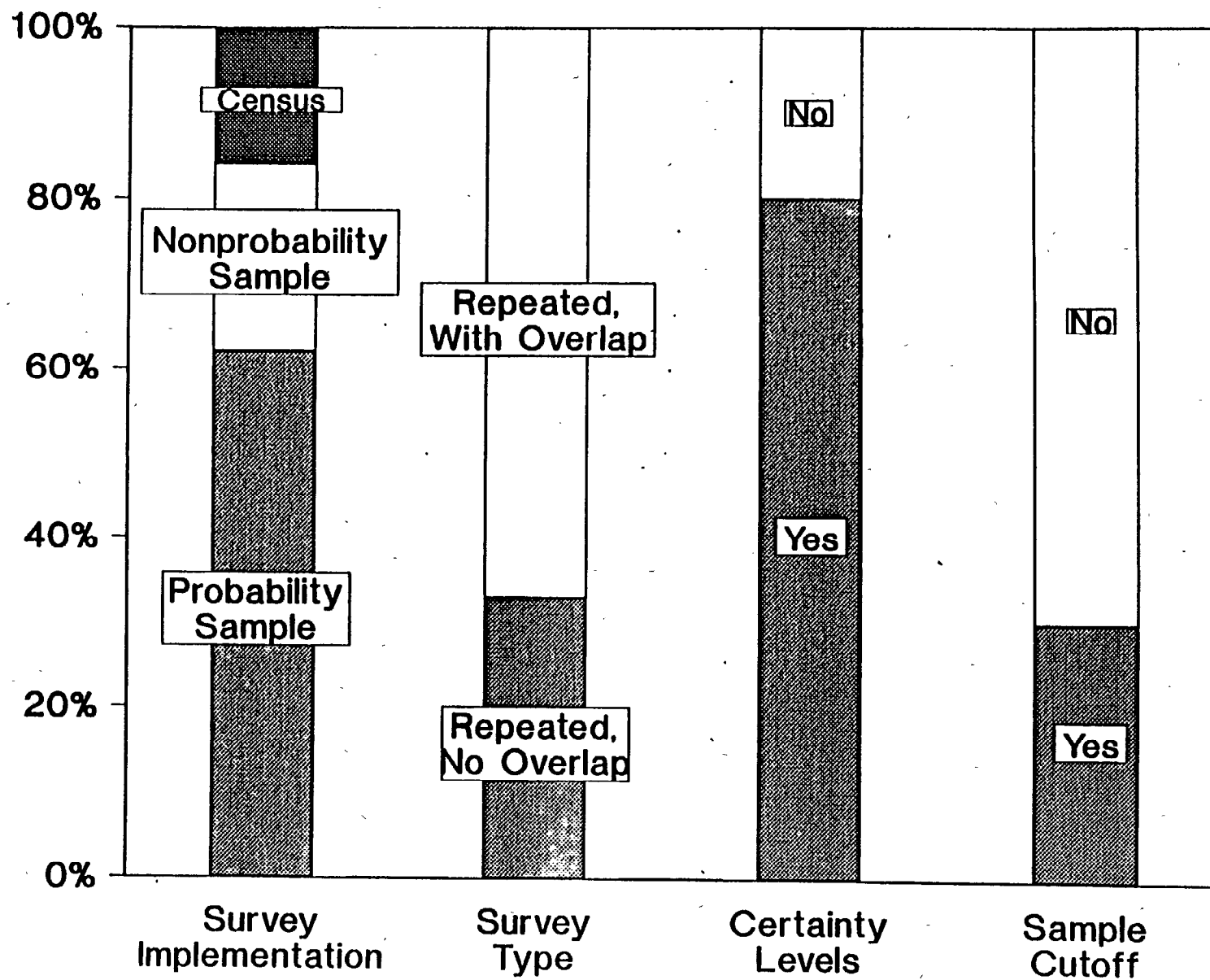
# SAMPLE DESIGN

Figure 2a



# SAMPLE DESIGN

Figure 2b



## **D. ESTIMATION**

### **1. BACKGROUND**

Without a measurement for the complete population of interest, a survey practitioner is forced to make inferences about the population based on sample estimates. The previous section discussed various areas to be considered in the actual selection of the sample. This section deals with how results from the sample are used to make estimates. There are several commonly used estimator types. The choice among estimators usually depends on the sample design itself and on the resources available to the agency for computing them. Before choosing a particular type of estimator, several things need to be considered. These considerations are usually made as a package at the time the sample is designed. For example, how was the sample selected? Was it a probability design or some nonprobability sample? What types of estimates, levels or changes, are desired? Is the survey going to be a one-time survey or will it be repeated several times? How many related items are to be measured? Are these items correlated with one another? Is there any known auxiliary information that can be used to improve the accuracy and precision of the estimates?

### **2. COMMONLY USED ESTIMATORS**

This section will discuss four commonly used estimators. Four areas for each estimator will be addressed. The areas include: What is the estimator? How is the estimator applied? Under what conditions should the estimator be used? What are the major advantages and disadvantages of its use?

#### **a) Direct Expansion Estimator**

This estimator applies some weighting or inflation factor to each sampled establishment. The inflation factor used is generally the inverse of the probability of selection of the establishment. For example, suppose a sample of 100 retail establishments has been selected at random from a population of 1,000 such establishments in a city. If simple random sampling without replacement has been used in the selection process, then each establishment will have 100/1,000 chance of selection into the sample. That is, the probability of selection of each establishment is 1/10. The Direct Expansion (Horvitz-Thompson) estimator can be used to estimate total sales for the city by multiplying the sales of each sampled establishment by the reciprocal of its probability of selection. In this example the direct expansion weight for establishment  $i$  ( $W_i$ ) is 10. The estimator is of the form:

$$\hat{Y} = \sum_{i=1}^n W_i Y_i$$

where

$\hat{Y}$  is the estimated total,  
 $W_i$  is the weight applied to sample unit  $i$ , and  
 $Y_i$  is the value reported by sample unit  $i$ .

The weights used in the Direct Expansion estimator do not need to be the same for each sampled unit. If, in the selection of the sample, some different probability of selection was assigned to different units, then the weight used in this estimator for each unit is the inverse of the probability of selection for that unit.

This estimator can be used in most simple probability designs. It is often used in establishment surveys since many establishment surveys are single-stage highly-stratified designs. This estimator can be used in cases with a random sample of units within strata with stratum weights of  $N_j/n_j$ , to be applied to each sampled unit in the  $j$ th stratum. In this case  $N_j$  is the number of population units and  $n_j$  is the number of selected units in the  $j$ th stratum. It can also be used in conjunction with a probability proportionate to size sample design with establishment weights being inversely proportional to the probability of selection. This estimator does not use any auxiliary information not used in the actual sample selection, but it can be used as the basis for other estimates which do use this information.

The advantages of the Direct Expansion estimator are that it is operationally simple, it is unbiased and its variance estimator has a linear form. Its major disadvantage is that it may not be a very efficient estimator.

#### b) Ratio Estimator

A second commonly used estimator is the ratio estimator. This estimator is used when the researcher has some additional information about the population of interest, such as a measurement of the variable of interest for some other period of time or perhaps the population value for some related variable. The ratio estimate utilizes this information to improve the predictive ability of the sample. For example, suppose one is interested in estimating total shipments for some manufacturing industry. A sample of establishments from this industry has been selected and data collected from each one. The shipments for each establishment in the sample in the previous census year is known from historical records. The shipments of the entire industry in that census year is also known. This information can be used to estimate the shipments of the entire industry in the current year.

Let

$\hat{Y}$  be the Direct Expansion estimator of the current year shipments,  
 $\hat{X}$  be the Direct Expansion estimator of the last census year shipments,  
 $X$  be the complete enumeration total for the census year shipments.

The ratio estimator  $\hat{Y}_R$  of the current year shipments is of the form:

$$\hat{Y}_R = \frac{\hat{Y}}{\hat{X}} X = \frac{\sum_{i=1}^n w_i Y_i}{\sum_{i=1}^n w_i X_i} X$$

In this example, when the variable  $Y$  (current year shipments) and  $X$  (census year shipments) are at least moderately positively correlated, the ratio estimator is an improvement over the simple Direct Expansion estimator.

Ratio estimation is often used in establishment surveys. Ratio estimation is particularly useful when the variables in the survey to be measured are correlated or when auxiliary information exists with some known total to adjust the estimates. To be effective, a plot of the  $X$  and  $Y$  variables should go through the origin or nearly so, and a positive correlation should exist. When this condition exists, gains in both accuracy and efficiency of the estimates can be realized. The ratio estimator is subject to a bias which arises from its nonlinear form. The size of the bias is a function of the sample size (small sample sizes are more subject to bias than larger sample sizes).

One additional problem faced by a researcher considering the use of ratio estimation is whether to use separate or combined estimates. That is, are ratio estimates formed separately for each sampling stratum and then summed across or are ratio estimates formed for all the strata combined? Cochran (1977) gives more detail on areas to consider in making this choice, with the sample size within the strata and the degree of correlation across the strata being the primary considerations.

### c) Link-Relative Estimator

When the primary interest is one of estimating period-to-period change, sometimes one may consider the use of the link-relative or link-change estimator. This estimator is similar in many ways to the ratio estimator. It is commonly used when poor levels of response and limited ability to impute make the use of a strict Direct Expansion estimator for the numerator and denominator of the ratio impractical. This estimator uses only the reported values of  $Y_i$  and  $X_i$  and may or may not include weights. It is used mostly to carry forward previous benchmark totals. For example, suppose the total ending inventories for establishments in a particular Standard Industrial Classification (SIC) code are known at the

end of the calendar year. A measure of how this value changes from month to month during the coming year is desired. The sample that has been selected is a cutoff sample representing some convenient group of establishments in the SIC code. Because of the nonrandom nature of the sample, stand alone estimates of monthly totals are not possible. However, if one is willing to assume that the month-to-month movements of the reporting establishments is adequate to measure the month-to-month movement of the universe as a whole, then a link-relative estimate may be used. The link relative estimate is of the form:

$$\hat{Y}_{L(i)} = \frac{\hat{Y}}{\hat{X}} \hat{Y}_{L(i-1)} = \frac{\sum_{j=1}^k w_j Y_j}{\sum_{j=1}^k w_j X_j} \hat{Y}_{L(i-1)}$$

where

$w_j$  is the inverse of the probability of selection of establishment  $j$ . (Often the weights are not used for link-relative estimator due to the nonprobability nature of the sample. In this case  $w_j=1$  for all establishments),

$\hat{Y}_{L(i)}$  is the link-relative estimate for month  $i$ ,  
 $\hat{Y}$  is the weighted total for the  $k$  reporting establishments in month  $i$ ,  
 $\hat{X}$  is the weighted total for the  $k$  reporting establishments in month  $(i-1)$ , and  
 $\hat{Y}_{L(i-1)}$  is the link-relative estimate for month  $(i-1)$ .

The link-relative estimator is biased. If the assumption that the responding establishments are representative of the universe is not true, estimates formed using this procedure are biased. In practice the bias can be severe. A common use of this estimator involves measuring change for very large establishments only and then assuming that the changes are reflective of the small establishments as well.

#### d) Unweighted Estimator

This estimator is used less frequently. Occasionally one is called upon to measure a highly skewed distribution, a cutoff of the largest units is selected and only those who report are tabulated. Typically the estimates are used to show relationships but they understate the true levels. Usually when this type of estimator is used, some attempt is made to indicate the degree of coverage the given sample has for the universe. For some establishment surveys, particularly establishments in manufacturing, the use of an unweighted sample benchmarked to control totals can be useful. This estimator is always biased even for trends but the cost and operational simplicity may cause it to be considered.

#### e) Estimation Techniques for Cutoff Samples

A number of establishment surveys employ a form of cutoff sampling in which no units are selected below a specified size. One cutoff design is

not actually cutoff sampling but rather a redefinition of the target population. In these cases the target population has been defined to be only units in the population with at least a specified size. Some surveys purport to be covering all establishments but just impute for units not given a chance of selection. Imputation may be either explicit or implicit. Explicit imputation methods typically use administrative data for the missing establishments as proxy for survey data. This is statistically sound as long as the concept being measured is identical in both data sources. Implicit imputation uses data from larger establishments or historical data as proxy data for units not surveyed. This latter approach is clearly less desirable since no current direct information is used for the establishment being imputed. A combination of explicit and implicit imputation is not uncommon within one survey.

## **E. SAMPLING ERROR ESTIMATION**

### **1. BACKGROUND**

The standard measure of the accuracy of an estimator is its mean-squared error. The mean-squared error is defined to be the expected value of the squared difference between an estimator and the value it is trying to estimate (Cochran, 1977).

The mean-squared error is composed of two parts. One part is a sampling variance and the other is a square of the bias component. Estimation assumptions can result in sources of bias. While the bias squared may be the dominant piece of the total mean-squared error, it is very difficult and expensive to measure and in practice little quantitative information about it is available for establishment surveys.

The sampling variance, the uncertainty caused by the fact that data is collected from only a part of the universe, is often estimable from the sample data itself. However, estimates of this statistic are included in publications of the data for only about half of the Federal establishment surveys.

Sampling variances are computed for roughly three-quarters of the establishment sample surveys of the Federal government. Sampling variances are used to quantify the accuracy of estimates and to confirm the sample design hypothesis. They are also used by some agencies as standards for what can and cannot be highlighted in press releases or in the narrative accompanying publications. Analysts often use these estimates to aid them in interpreting agency statistics.

### **2. COMMON APPROACHES TO VARIANCE ESTIMATION**

There are numerous different approaches to the calculation of sampling variances. Wolter (1985) is devoted entirely to the estimation of variances. The text provides an exhaustive treatment of most of the currently used methods of variance estimation as well as some rationale for choosing among them. This paper will briefly discuss only a few of the more commonly used approaches.

#### **a) Design-Based Variances**

The actual sampling variance of a survey statistic is a function of the form of the statistic and of the nature of the sample design. The variance of a statistic  $Y$  is defined as

$$\text{VAR}(\hat{Y}) = E(\hat{Y} - E\hat{Y})^2$$

For simple sample designs with simple linear estimators, it is often possible to directly compute the estimates of  $\text{VAR}(\hat{Y})$  from the sample data. These design-based estimates of variance depend on how the sample was selected and specific formulas for their computation can be found in most standard sampling texts (Cochran, 1977 and Wolter, 1985).

This direct approach to variance estimation is desirable and should be used whenever possible. Unfortunately, in practice, the type of estimator used may be so so complex that it is impossible to derive a direct design-based variance formula.

#### b) Replication Estimators of Sampling Variance

There are instances of highly complex sample designs in which an accurate estimate of sampling variance cannot be obtained from a single sample unless certain generalizing assumptions are made concerning the universe. This is generally due to the extremely complicated nature of the variance formulas. Variance estimates based on replicates, however, can be used to simulate the effects of all aspects of the sample that vary from replicate to replicate, and this greatly increases the computational efficiency of sample variance estimation.

Besides aiding sample variance estimation, there are other factors that lead survey practitioners to use replicate estimates. The ordinary Taylor series approximation for obtaining the estimated variances of ratio estimates, even for simple random sampling, provides an estimate even though biased. Sometimes drawing a number of independent samples, computing a ratio estimate for each sample and then averaging these ratio estimates for the final estimate is used. A valid estimate of sampling variance can then be developed from the replicated values of the estimate.

#### c) Random Groups

This method of variance estimation is applied by taking  $k$  disjoint subsamples from the sample and forming the estimates of  $\Theta_\delta$  ( $\delta = 1, 2, \dots, k$ ) for each subsample. The random group estimator of the parameter  $\Theta$  is then

$$\hat{\Theta} = \frac{1}{k} \sum_{\delta=1}^k \hat{\Theta}_\delta$$

where its variance is estimated as

$$\text{Var } \hat{\Theta} = \frac{1}{k(k-1)} \sum_{\delta=1}^k (\hat{\Theta}_\delta - \hat{\Theta})^2$$

#### d) Generalized Variances

Suppose a simple mathematical relationship or model exists between the variance of a survey estimator and the expected value of the estimator. Then if the parameters of the model can be estimated from past data or from a small subset of the survey items, variance estimates can be produced for all survey items simply by evaluating the model at the survey estimates rather than by direct computations. This method of variance estimation is called the method of Generalized Variance Functions (GVF).

In general, GVF's are useful for surveys that publish a large number of different statistics for several different subgroups. When the number of published estimates is manageable, we generally prefer direct measures of the variance. The primary reasons for considering GVF's include:

1. Even with modern computers the cost of a direct computation of variance for each one of many statistics may be excessive.
2. Even if the cost is affordable the problems of publishing all variance estimates may be unmanageable.
3. It may not be possible in advance to anticipate all the types of statistics for which variances will ultimately be desired.

The difficulty of using this procedure is of course in selecting and fitting the correct model. This is not as easy as it sounds, and hence this method is not widely used for establishment surveys.

#### **e) Taylor Series Methods**

In surveys it is desirable to develop estimators that are not linear. Examples of these types of estimators include ratios, differences in ratios, correlation coefficients, regression coefficients, etc. Exact expressions for the variance of these estimates are not usually available. Even simple unbiased estimators of the variance may be lacking.

One useful method of estimating the variance of a nonlinear estimator is to approximate the estimator by a linear function. Once this is done one can develop an estimator for the variance of the linear approximation and use it as an estimator for the variance of the nonlinear one. This procedure is biased but is typically consistent. The validity of this procedure relies on the use of the Taylor Series or binomial series expansions and hence the name Taylor Series Variance Methods.

### **3. FACTORS AFFECTING THE USE OF VARIANCES IN ESTABLISHMENT SURVEYS**

Establishment surveys conducted within the government cover a broad range of sample designs and variance estimators. Probability samples are generally preferred, but are not uniformly used. The reasons given for not using probability designs vary, but resource constraints seem to be a common element in all of them. The cost of ensuring coverage and maintaining the representative nature of the survey is not inconsequential. Even when a good probability design is selected and maintained, it is likely that the nonresponse pattern will not be random and will result in biases in the estimates. The two main motivations for probability design are the representative nature of the sample and the ability to compute variances from probability samples. The extent to which variances are actually computed varies both as to frequency and as to the level of detail. Reasons for not computing and/or not publishing variance estimates for surveys relate to the cost both in time and computer resources of computing variances and to the perceived lack of use of such measures. In order to accurately compute variances, additional data files need to be maintained and utilized. Timing for establishment surveys is critical

and the delay needed to compute variances is sometimes viewed as too great a price to pay.

For some surveys, particularly economic indicator surveys, where the period-to-period trend is judged to be the primary measure of interest, often nonprobability designs are used. They are generally simpler to use and maintain and the biases associated with incomplete coverage of the universe are not as serious in the measurement of change. For these non-probability surveys, variances are not computed. For some surveys, general measures of mean square errors based on levels of revisions are computed to give the user a rough idea of sample variability.

The general consensus is that a well maintained probability sample design with frequently computed and published variance estimates is the ideal standard. Lack of resources to devote to the work of maintaining the samples and computing the variances results in many designs not meeting these standards.

#### 4. SUMMARY PROFILE

(See Figure 3.) Information on estimation and variance estimation was collected as part of the profile of survey practices. The Economic Censuses were excluded from this part of the analysis. Figure 3 illustrates some interesting characteristics of the measured surveys.

Most survey estimates were either Direct Expansion or ratio type estimates. The link relative form of estimates was used for roughly 15 percent of the surveys with around 10 percent of the surveys reporting some other type of estimation. Generally surveys measuring indexes or month-to-month changes were more likely to use a link-relative or other form of estimator. The more traditional estimates of totals were generated by expansion or ratio type estimators.

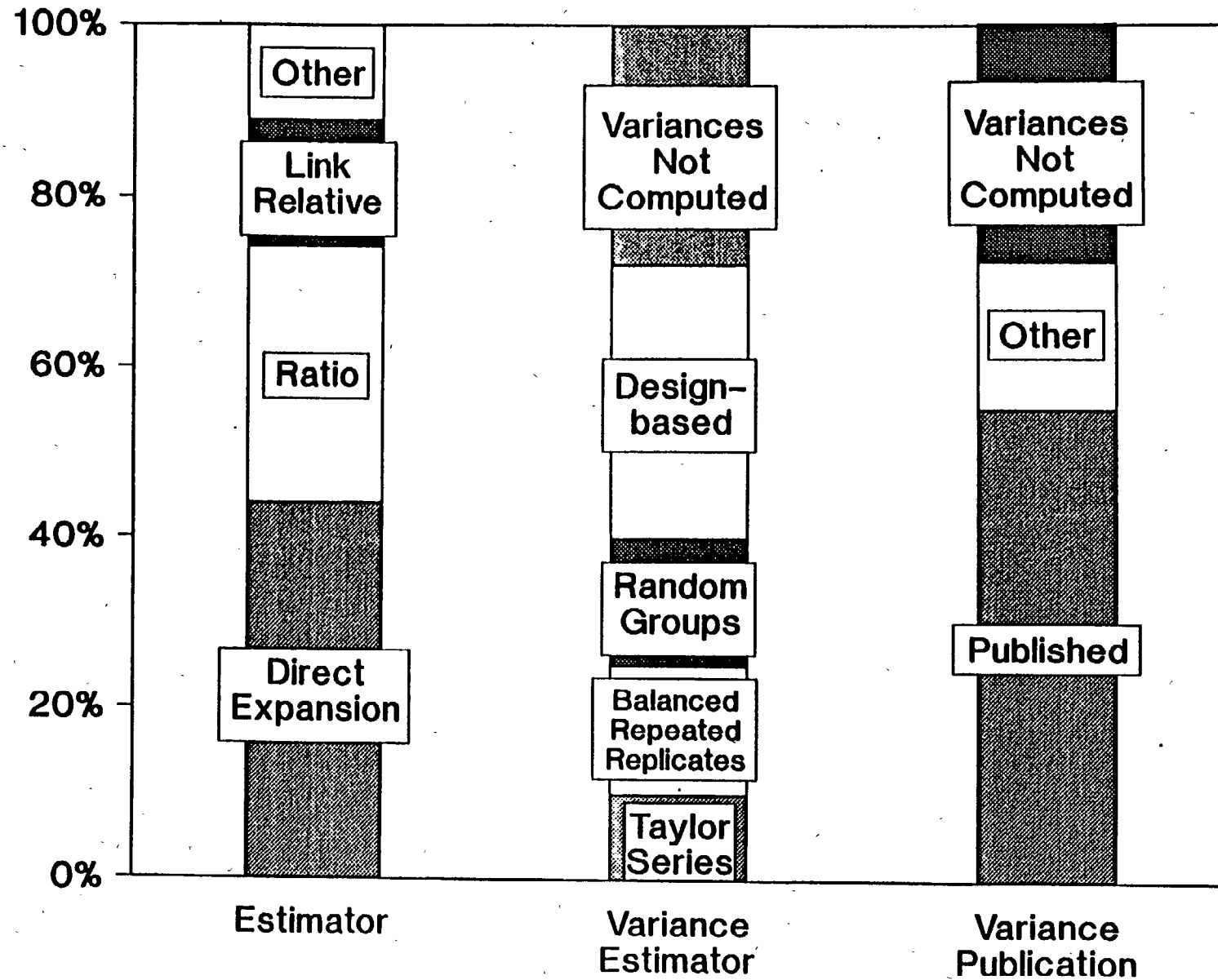
In the area of variance estimation several interesting findings are apparent. Slightly over one quarter of the sample surveys do not compute variances at all, even for internal purposes. Approximately one-third of the sample surveys used a design-based variance formula which varied from survey to survey due to the nature of the sample design. The remaining sample surveys used a replicate or Taylor series method of variance estimation.

The sample surveys are classified by whether or not the variances were included in the publications. Almost half of the sample surveys covered do not publish variances. This seems unusually high and marks a major difference between household and economic surveys.

The distribution of surveys not showing variances did not seem to be confined to one or a few agencies but in general when link-relative or other nonstandard estimation was employed the variances were not published. A second theme not specifically shown in the figure but frequently mentioned was the perception on the part of survey analysts that their users neither know nor understand what variances are. This view of the relative unimportance of measures of reliability may well have contributed to the high percentage of surveys not publishing variances.

# ESTIMATION

Figure 3



## CHAPTER IV. SURVEY METHODS AND OPERATIONS

### A. INTRODUCTION

#### 1. BASIC CONCEPTS

This chapter focuses on the errors which arise during the specifications for and the conduct of establishment surveys. The errors which occur during these operations are called nonsampling errors. Commonly known examples of nonsampling errors include incomplete sampling frames, non-response and keypunching errors. A survey design consists of a large number of methods and operations. Each method or operation is a potential contributor to nonsampling error. Such variety of nonsampling error sources leads survey researchers to believe that nonsampling errors may far exceed sampling error. Establishment surveys are no exception, which makes understanding nonsampling error essential for understanding establishment survey results. The primary objectives of this chapter are to outline major categories of nonsampling errors in establishment surveys, to identify some of the diverse sources of error in each category, and to provide insight into strategies to detect, measure, and control these errors. The error categories discussed are specification, coverage, response, nonresponse, and processing errors.

#### 2. ERROR MEASUREMENT

The importance of nonsampling errors has led to the concept of "total survey design" in which measurement and control of both sampling and nonsampling error are given consideration during the initial design of the sampling plan. The diversity of nonsampling error sources combined with the numerous complex survey designs used in establishment surveys makes it difficult to address all the possible designs for nonsampling error evaluation. Most survey researchers agree that a measurement of the total bias should be obtained if it is feasible. Unfortunately, the true value is needed to measure total bias, and for many establishment survey data items the true value is either impossible or too costly to obtain. When this is the case, procedures which evaluate individual sources of nonsampling error are recommended. Often an error profile is developed to guide the survey researcher toward the specific sources of nonsampling error which should be studied. These special studies often assume a particular model structure of the errors and are designed to measure parameters of the model. Validation studies and interpenetrating samples are common methods used to study nonsampling errors. Several specific examples are given in this chapter.

As an aid to understanding the impact of nonsampling errors, techniques to directly or indirectly measure nonsampling error will be discussed for each of the nonsampling error categories which this chapter will review. Direct measurement techniques typically provide an estimate of the bias or variable error resulting from an error source; for example, a post-survey followup of a sample of nonrespondents. Indirect measurement techniques typically provide an indication of the potential for bias or variance resulting from an error source, but not an estimate of the bias or variable error; for example, the nonresponse rate.

## **B. SPECIFICATION ERROR**

### **1. DEFINITION OF SPECIFICATION ERROR**

Specification error is the error that occurs at the planning stage of a survey because data specification is inadequate and/or inconsistent with respect to the objectives of the survey. In an economic survey, it is often the difference between the quantity intended to be measured, such as the price or volume of a good, and the data collector's ability to obtain this measure. Specification error can result simply from poorly worded questionnaires and survey instructions or may reflect the difficulty of measuring abstract concepts.

#### **Example**

A type of specification error that frequently arises in energy-related surveys relates to the concept of consumption. Data on actual consumption of energy is difficult and costly to collect because most energy producers do not keep records on the final consumption of their products. For this reason, respondents to energy-related surveys may be asked to report on deliveries, products supplied, or sales. Because these data do not measure energy consumption directly, their use as a proxy for consumption data introduces some degree of error into energy consumption statistics.

### **2. SOURCES OF SPECIFICATION ERROR**

Three sources of specification error are discussed in this section: (1) inadequately specified uses and needs, (2) inadequately specified concepts, and (3) inadequately specified data elements.

#### **Inadequately Specified Uses and Needs**

Behind every survey is some need for the data. It may be to report on economic conditions, support a legislative program, or allocate Federal funds. Whatever it is, the sponsor of a survey has a use for the data. When the uses and needs documented for a survey do not correspond to the actual uses and needs for the data, specification error occurs.

There are several causes for inadequately specified uses and needs. These include (1) poorly stated uses and needs by the sponsor, (2) changing uses and needs over time, and (3) the population of inference not corresponding to the population surveyed.

**Poorly stated uses and needs**--The sponsor of a survey is responsible for specifying the uses of the data. This often requires the sponsor to conduct a special study or data needs assessment to identify data uses. If the uses are poorly defined and not specific, then it will be difficult to correctly specify what data are to be collected. This will result in specification error biasing the data from the outset.

The data collector is also responsible for specifying the needs and uses of the data. Very often the data collector has experience in meeting a specific set of sponsor and user needs, and knows what kind of data are needed to meet program requirements.

Finally, potential users of the data must be consulted as to their needs for the data. When a Federal agency sponsors a survey, a notice is published in the Federal Register asking for comments. Not only do potential respondents make comments, but potential users of the data often comment on whether the data will meet their needs. When the needs of other users do not coincide with those of the sponsor, even careful data specification may not satisfy all parties. While not an error in the traditional sense, this can be classified as specification error since when one party uses data collected for the other's needs, it will not be properly specified.

**Changing uses and needs**--Data needs change over time; consequently they must be reexamined on occasion. Even if the needs are clearly and unambiguously stated when the survey was undertaken, periodic review of data requirements is necessary to take into account changes in business and industry, changes in legislation, and changes in user requirements which will affect what data need to be collected.

**Population of interest not same as population surveyed**--Specification error can occur when the survey respondents are not the same as the population for which the estimates are needed. This can occur when a survey is created for one sponsor and questions are added by another sponsor to save costs associated with creating an entirely new data collection. It can also occur when the population of interest is not obtainable because of frame deficiencies. In these cases the surrogate population is surveyed and estimates are produced. The surrogate population may not be able to answer the questions accurately or in the same way as the "real" population would have. This may not be an error in the strict sense of the word, but it would result in the estimated data measuring something different from what was intended by the survey sponsor.

#### **Inadequately Specified Concepts**

Once a need has been identified, it must be stated as a measurable concept. Specification error reflects the extent to which concepts defined for a survey do not reflect the primary uses and needs for the survey data. This may either be the result of using concepts that are poorly defined or of using existing concepts that do not fit the need.

**Poorly defined concepts**--Survey concepts must be unambiguously and carefully worded. Suppose an agency needs to know the amount of coal produced annually in the United States. It is critical to consider at the outset whether the types of coal produced--lignite, bituminous, and anthracite--need to be distinguished and whether production is defined as what is "dug out" of the ground or what has been cleaned and prepared for shipment.

**Using an existing concept that does not really fit**--A poorly specified data need is as likely to cause specification error as a poorly defined concept. Consider again, for example, the problem of determining energy consumption. Assume the sponsor or data user is interested in how much

energy is used by a particular type of consumer, such as an industrial plant or commercial establishment, at the State level. The concept of interest here is end-use consumption. This is most accurately measured by going to the end user. However, this would be very costly and time-consuming because of the large number of end users. Instead, a surrogate measure, such as products supplied, may be used because there are far fewer energy suppliers than consumers and the data are more easily disaggregated to the State level. Nevertheless, inaccuracies may result since supplied energy can be stored for later use or may be resold to other consumers. Thus using the concept of "product supplied" in lieu of measuring end-use consumption may well introduce error into the estimates. This points up the need for surveys that directly measure a phenomenon. In the case of end-use consumption, triennial consumption surveys are conducted to measure energy use from the consumer. Although more costly and time consuming, they serve many important functions including that of a benchmark against which to measure the adequacy of surrogate measures.

A related notion is one where a measure is adequate for one purpose but is flawed for another. Consider the example of stocks such as coal in a pile at a utility or crude oil in a storage tank at a refinery. In both cases what is at the bottom of the pile or tank is not usable. If the need is to identify month-to-month changes, then measuring stocks as a total volume is adequate. If, however, the need is a measure of quantities on hand in case of a supply disruption, then the measure is not adequate.

#### **Inadequately Specified Data Elements**

Data elements may be defined on the questionnaire in such a way that they do not accurately reflect the survey's intention. This is another source of specification error. Inadequate specification of data elements may result from (1) ambiguous definitions, (2) elements that do not fully reflect the survey concepts, (3) use of proxy data due to unavailability of primary data, and (4) poorly worded questions.

**Ambiguous definitions**--Ambiguous definitions may result in respondents reporting different data than is intended by the sponsor of the survey. For example, in a survey of crude oil production, it would be important to carefully define the term "crude oil." Otherwise, respondents would be left guessing whether, for example, to include lease condensate, a natural gas liquid recovered from gas-well gas, in their crude oil production figures. Because lease condensate is generally blended with crude oil for refining, some producers might automatically include it in reported volumes of crude oil production. Others might not include it in the reported volumes, or might report it separately. Thus if crude oil were not clearly defined in the data collection instrument, respondents would likely use varying definitions in reporting production figures. Precise specification, then, is the key to achieving consistent responses that measure the intended concept accurately.

**Elements not reflecting survey concepts**--All research entails describing or analyzing certain theoretical concepts. In establishment surveys it might be the money flow among federally chartered banks, the supply of petroleum products, or the behavior of producer prices in the economy.

Before data can be collected and analyzed, these concepts must be reduced to specific, empirical indicators. The data collector must specify observations that may be taken as indicators of the attributes of a given concept. An operational definition must be created that will measure that concept. The process is complicated in establishment surveys because economic statistics are usually byproducts of other business or government activities and have to be collected as part of that process. Thus data collectors often lack control over what is collected, how it is defined, and how closely the definition conforms to the concept being measured.

Moreover, when several variables are used to create a composite measure, such as a producer price index, the analyst has created a measure of an abstract concept that does not exist in any real economic sense. Error can then result not only from error in the individual variables, but can be compounded when these statistics are combined.

Proxy data requested due to unavailable primary data--Even where concepts are clearly defined, respondents may be unable to supply the requested data because the data are not available. Another energy-related example involves the disaggregation of natural gas supplied by end-use sector. Generally, utilities keep track of gas supplied by rate class--industrial, commercial, and residential. However, these classes are determined not by the actual function of the energy consumer, but by the flow rate or amount of energy consumed. This is also how the public utility commissions determine utility rates. Thus master-metered apartment buildings may get billed at the commercial rate rather than at the residential rate. As a result, the utility may be unable to provide accurate information broken down by end-use sector even when the sectors are clearly defined. Moreover, because of the great differences in rate classes in different States, inconsistencies between States can lead to errors in the national figures that are hard to detect and quantify.

Questionnaire wording, definitions, classification, or instructions--Once an operational definition has been specified, a survey instrument is constructed, questions are formulated, terms are defined, and instructions for completing the questionnaire are written. Ambiguous questions, questions without unique answers, and unclear instructions all cause response errors. Misclassification may occur when respondents are asked to report familiar data in ways that are unfamiliar to them or in inconsistent ways. For example, companies reporting on imported petroleum products are asked to classify commodities one way for the U.S. Custom Service and another way for the Department of Energy. Both schemes have legitimate conceptual foundations, but the disparity in definitions causes difficulty both to the respondents and to the data collectors.

Respondent classification is another major source of specification error, particularly when multifunctional conglomerates are assigned SIC codes, or when parent/subsidiary relationships have to be untangled. Moreover, the risk of double counting increases when data are aggregated from several surveys in which the rules for classification are unclear or inconsistent.

### **3. CONTROL OF SPECIFICATION ERROR**

Control of specification error relies on the tenets of good questionnaire design as well as some of the techniques used in its measurement (which are discussed in the following section). These control mechanisms include (1) requirements reviews, (2) industry consultations, (3) expert review panels, (4) cognitive studies, and (5) pretests.

#### **Requirements Reviews**

A requirements review determines what data in a subject matter area are needed. Potential data users and analysts are contacted to find out if new data are required and how these data would be used. Data that are currently being collected are evaluated to determine if they meet users' analytical needs. If not, this may suggest that the wrong data are being collected. This can frequently be remedied by changing some of the definitions used in the survey in lieu of collecting new data. The steps involved in conducting a requirements review are: (1) assembling available background information on the phenomenon to be measured, (2) developing a description of the phenomenon, (3) researching and formalizing the evidence from which to infer information requirements, (4) generating a matrix of data requirements with relationships mapped to the need for the information, (5) developing a rationale for selecting the required data, (6) developing the "justified" data requirements by applying the rationale to the data requirements matrix, and (7) identifying new data elements or changes in existing elements that need to be implemented.

#### **Industry Consultations**

Whenever a new data collection instrument or changes to an existing instrument are proposed, the agency sponsoring the survey should discuss the proposed instrument with those who will be supplying the data. This can be done through discussions with trade associations and industry representatives as well as directly with potential respondents. Operational definitions can be discussed, recordkeeping practices reviewed, and data collection methodology explained. Allowing potential respondents to provide input into the data specification process helps ensure that the survey elements will be properly specified.

#### **Expert Review Panels**

Sometimes it is useful to convene a panel of experts in the subject matter area of the survey to review the specification of data. The panel is usually assigned a specific task--such as a review of definitions of petroleum products or of unemployment. The panel's recommendations help ensure that questionnaires and instructions meet the stated objectives of the study and measure what they purport to measure.

#### **Cognitive Studies**

Cognitive studies, which are discussed in more detail in the following section on measurement of specification error, can be used both to measure specification error and to control it. In the process of measuring an error, the causes for that error are often uncovered. Steps can then be

taken to control the problem by revising the definitions, changing the wording of the questionnaire, or modifying the instructions.

#### **Questionnaire Pretests**

Pretesting questionnaires is another activity essential for both measuring and controlling specification error. Identifying and resolving problems with the survey instrument before it is used in a full-scale data collection reduces specification error in the final study.

#### **4. MEASUREMENT OF SPECIFICATION ERROR**

Specification error can be measured either directly or indirectly. Direct measurement of the error involves comparing the data value against some benchmark known to be true and accurate. The benchmark need not be the same as the data value, but the difference between the two should be a known constant. A method of direct measurement is records check surveys.

Indirect measurement techniques identify discrepancies or possible errors in the data. These techniques establish the existence of an error, often providing a qualitative description of it. An indirect measure can be quantified, but in the absence of a benchmark or "true" value against which to measure its magnitude and direction, the measure is only indirect. Indirect measures included cognitive studies, questionnaire pretests, and comparisons to independent estimates.

##### **Records Check Studies**

Specification error can be measured directly by checking survey responses against administrative records. This can involve auditing a company's books or matching survey responses against tax records or licensing information. Administrative records are not always available, however, because of privacy restrictions. When reviewing administrative records, it is important to determine whether definitions used in recordkeeping are the same as those used by the survey instrument. It is also important to determine whether there is an inherent bias in the recordkeeping because respondents over- or underreport for business or economic reasons.

##### **Cognitive Studies**

A cognitive study, or validation study, is an indirect approach to measuring specification error. It entails examining each stage of the data collection process from beginning to end to detect errors caused by improper operational definitions. This includes a review of data requirements, construction of the questionnaire and survey frame, data processing and editing procedures, nonresponse followup, and data aggregation and publication of results. Generally a site visit to selected respondents is the most useful way for identifying error associated with poor questionnaire design or disparate recordkeeping practices. Actually walking through the industrial or commercial process with the respondent is helpful. Seeing at what points the data are collected, how they are measured, and how they are used by the respondent will indicate whether the intended concepts are being accurately measured. In many respects this process is

similar to a pretest or pilot study, except that it is conducted after a survey is under way.

The disadvantage of cognitive studies is that they are very costly and labor intensive. Moreover, because the review concentrates on a very few respondents, it may be difficult to know whether the identified problems are widespread. This makes it difficult to quantify the magnitude of the errors discovered, even if it is possible to quantify the magnitude for that subset of the respondents.

### **Questionnaire Pretests**

Before a questionnaire is used in a study, it should be pretested and the results analyzed in the same way the actual data will be collected and analyzed. Many problems involving unclear definitions or the wording of questions and instructions will become apparent at this point.

### **Comparisons to Independent Estimates**

Another less costly technique for measuring specification error involves comparisons of data series. The data series in question is compared with similar, independent estimates. When the two estimates match up, both are usually presumed accurate. When the two estimates differ systematically, it is an indication that one of the estimates is biased. Sometimes the "true" value is considered bounded by the two estimates. If there is an indication of bias, one or more of the following procedures is instituted: (1) matching individual respondent records from the two data series, (2) contacting respondents, and (3) contacting the survey managers and data processing specialists to try to determine the source of the bias.

For example, as part of its annual assessment of data quality, the Energy Information Administration (EIA) compares its coal production data with similar data from other sources. In comparing EIA production data with information from the Mine Safety and Health Administration (MSHA), the MSHA data were found to be systematically lower than the comparable EIA data. The discrepancy ranged from 4.7 percent in 1978 to 2.6 percent in 1982. The comparisons were then disaggregated by type of coal, type of mine, and selected States to determine the possible causes for the discrepancies. It turned out that different definitions of clean versus raw coal accounted for some of the discrepancy in production figures.

## **5. SUMMARY PROFILE**

(See Figures 4 and 5.) In identifying procedures used by Federal statistical agencies to control specification error, the two most commonly used techniques employed were the requirements review and respondent consultation. This is not surprising given the requirements for forms clearance established by the Office of Management and Budget. A substantial number of agencies also have the questionnaires reviewed by expert panels. Surprisingly, relatively few surveys are pretested on a regular basis. Pretesting is done, however, when a survey is first started or if major modifications are made. Cognitive studies, on the other hand, which are expensive and time consuming are not often done, especially on a regular basis.

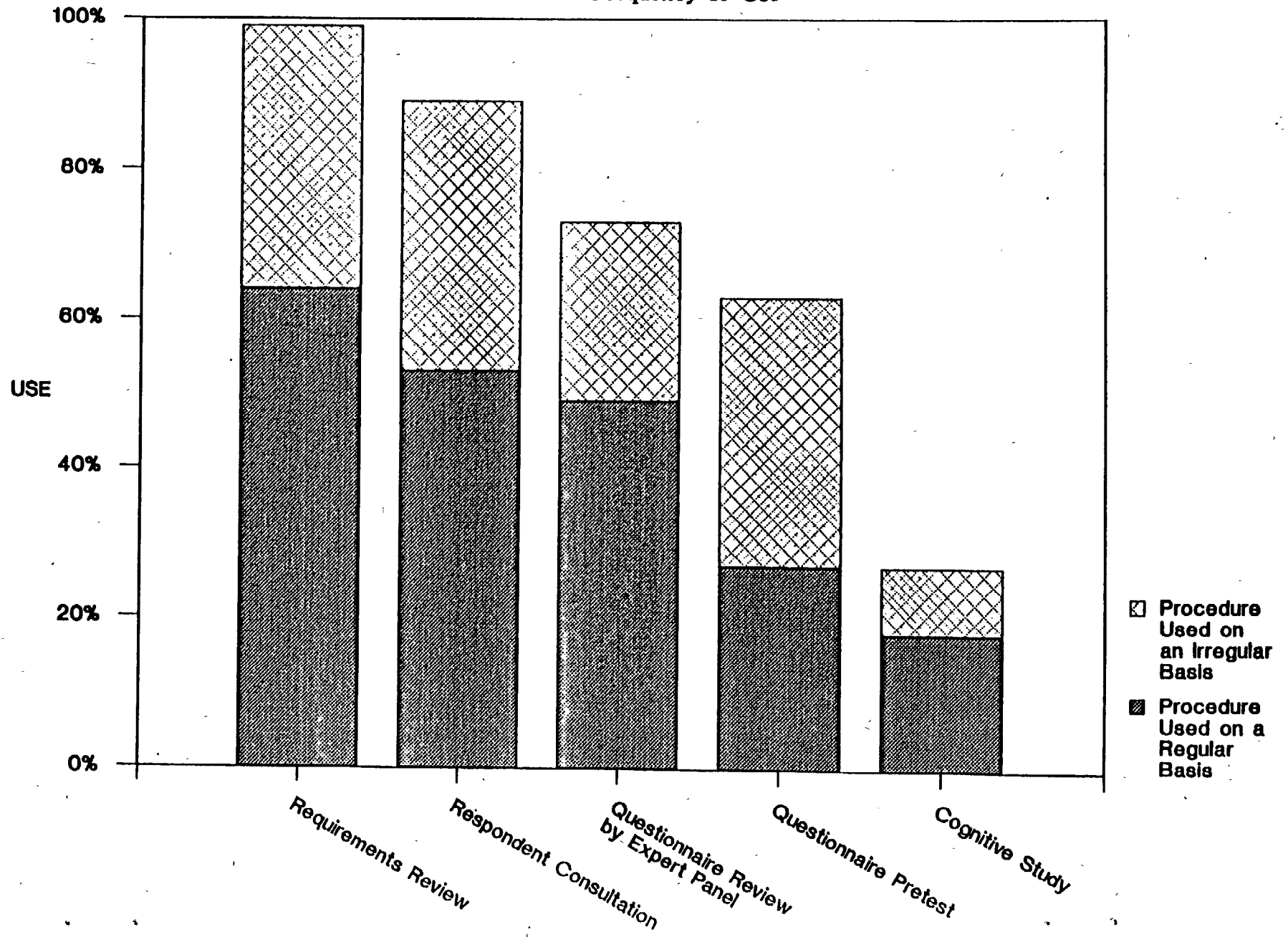
In general it appears that most of the agencies are taking steps to control specification error on the majority of their surveys. This is much less true when it comes to measuring specification error.

As Figure 5 shows, relatively little is done to measure specification error in establishment surveys. The most prevalent technique used to measure this source of error is comparison to independent estimates. It is the simplest and least expensive of the techniques and provides some quantitative measures of the direction and magnitude of the error. Relatively few surveys publish this comparative information. More should as it would be helpful to users of the data.

# SPECIFICATION ERROR CONTROL PROCEDURES

Extent and Frequency of Use

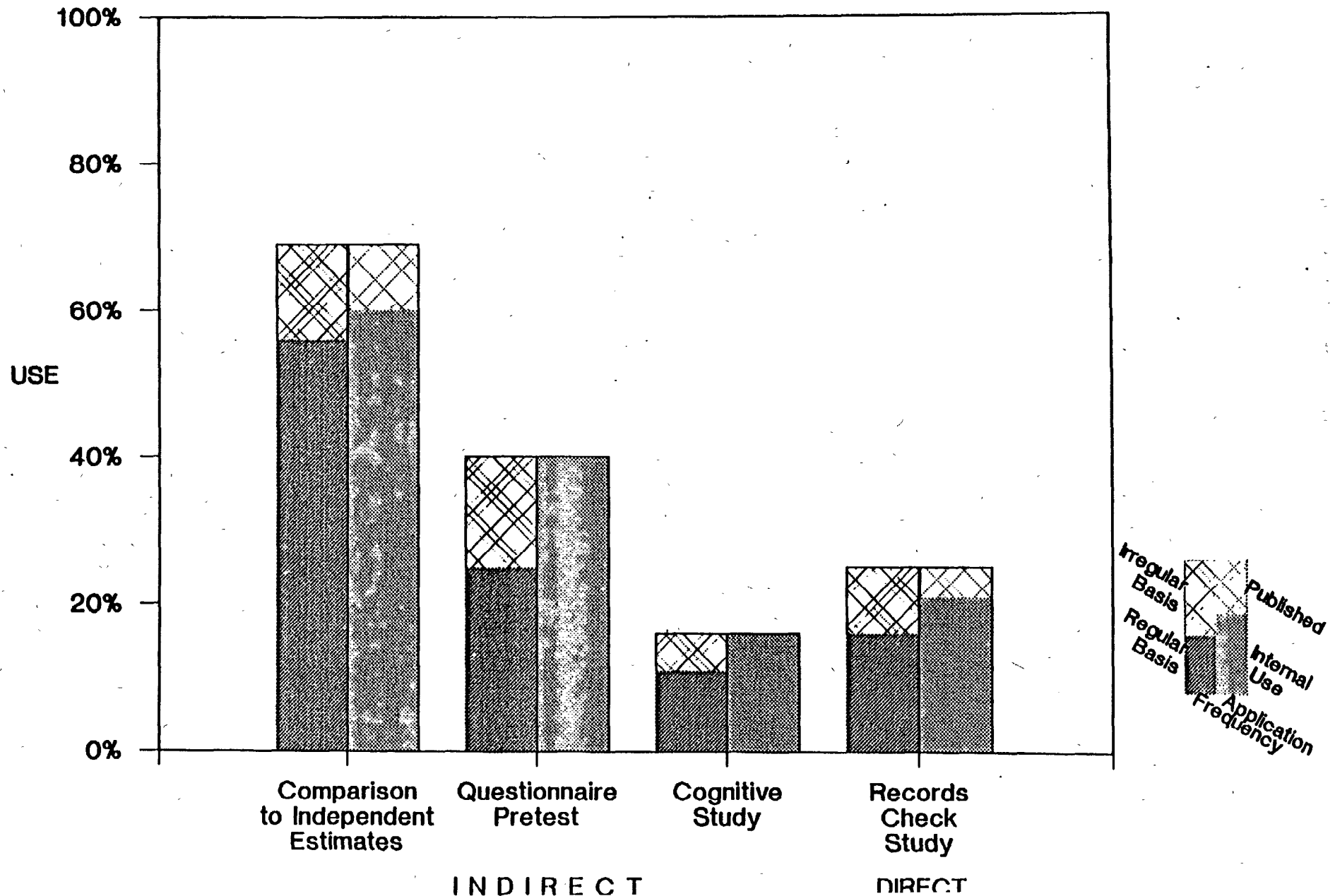
Figure 4



# SPECIFICATION ERROR MEASUREMENT TECHNIQUES

Frequency and Application of Use

Figure 5



## C. COVERAGE ERROR

### 1. DEFINITION OF COVERAGE ERROR

Coverage error, which includes both undercoverage and overcoverage, is defined as "the error in an estimate that results from (1) failure to include all units belonging to the defined population or failure to include specified units in the conduct of the survey (undercoverage), and (2) inclusion of some units erroneously either because of a defective frame or because of inclusion of unspecified units or inclusion of specified units more than once in the actual survey (overcoverage)" (Office of Federal Statistical Policy and Standards, 1978). Coverage errors are closely related to but clearly distinct from content errors, which are defined as the "errors of observation or objective measurement, of recording, of imputation, or of other processing which results in associating a wrong value of the characteristic with a specified unit" (Office of Federal Statistical Policy and Standards, 1978). Thus, an interviewer's failure to properly identify and hence to record data for what should be a selected unit is a coverage error. On the other hand, failure to pick up data for a properly selected unit (which results in an imputed value being assigned to the unit) is a content error. Content errors include response and nonresponse errors, both of which are discussed more fully elsewhere in this chapter.

### 2. SOURCES OF COVERAGE ERROR

While the definition divides coverage error into two major components--undercoverage and overcoverage--another important duality is implied within each of these: Coverage error shows up (1) in defective sampling frames and (2) as a result of defective processes associated with the selected sample. (Sampling frame, or stated simply, frame is used here to mean the collection of potential sampling units, either given explicitly as a list or implicitly in terms of well-defined procedures.) Thus coverage error results either because the frame does not properly represent the sampled population, or because the sample does not properly represent the frame. Note that, using the definitions of Cochran (1977), we are making a distinction between the sampled population, defined as the population to be sampled, and the target population, defined as the population about which information is wanted, if possible. Ideally, the sampled and target populations should coincide. However, cost or other practical considerations sometimes result in a lack of coincidence between the two. Consequently, the target population is sometimes modified to coincide with a workable sampled population.

Any difference between the sampled and target populations can contribute importantly to coverage error, especially where excessive compromise in the survey planning stage results in a sampled population which is too far removed from the target population. Since estimates based on data drawn from the sampled population apply properly only to the sampled population, interest in the target population dictates that the sampled population be as close as practicable to the target population. Nevertheless, in the following discussion of the sources, measurement, and control of coverage error, only deficiencies relative to the sampled population are included. Thus, when speaking of defective frames, only

those deficiencies are discussed which arise when the population which is sampled differs from the population intended to be sampled (the sampled population).

#### Coverage Error Source Categories

The two categories of coverage error--defective frames and defective processes associated with the selected sample--are discussed below.

**Defective Frames**--Defective frames are characterized by (1) deficiencies in meeting the requirement that every element of the sampled population belongs to one and only one sampling unit, (2) erroneous inclusion of units (including the wrong units or having duplicate units which belong in the frame), or (3) erroneous exclusion of sampling units. These problems can result from vague or unworkable definitions of the sampling units relative to the sampled population; improper procedures or processing in establishing and maintaining the frame; timing, which affects the updated-ness (agreement with the proper reference period) of the frame; or miscoding of sampling units. Erroneous inclusion (overcoverage) results from including duplicates and out-of-scope or out-of-business units. Erroneous exclusion of sampling units (undercoverage) results from failure to include the proper units or failing to account for birth (new) units. Misclassification of units, such as for SIC, geography, size class, or company structure can lead either to undercoverage or overcoverage.

Some frame problems cannot be overcome without expending significant resources. For example, most frames suffer from some degree of outdatedness. A monthly survey in which the frame and sample are updated quarterly, such as the Census Bureau's Monthly Wholesale Trade Survey (MWTS), does not have an up-to-date frame for at least two out of every three months--and this is over and above the lag time in getting new units on the list frame. Because the cost and processing difficulties preclude correcting for this frame error, the Census Bureau accounts for new units in its estimates by an imputation technique. The overall objective is to correct errors which can be corrected within resource limitations and thereby keep coverage error as low as is feasible. This time lag itself can be as much as 12 to 18 months after a business starts up. For example, the Social Security Administration (SSA) lists of EI numbers newly assigned by Internal Revenue Service (IRS) are given to the Census Bureau after SSA receives the EI application forms from IRS and codes them. Each processing step contributes to the lag.

**Defective Processes Associated with the Selected Sample**-- Coverage errors in which the selected sample does not correctly represent the frame may be the result of selected cases being inadvertently dropped from the sample or nonselected cases being added to the sample erroneously. Also, errors may be made in selecting the sample. Errors of this type are likely to occur when the sample is determined by interviewers in the field. In business area samples where the sampling units are geographic land segments, failure to properly identify the population units (business establishments of a particular type) is a common form of coverage error. Such errors may result from inadequate definitions or inadequately specified field or office procedures, outdated or otherwise incorrect maps of selected area sample units, or misapplication of the sampling or

canvassing rules by the interviewer. Failure to sample from an updated frame on a timely basis also results in a sample that is not representative of the sampled population. For other papers which discuss coverage concepts and issues, see Garrett et al. (1986) and United Nations (1982).

It is worth noting here that even where coverage of a total population is fairly good, serious problems may exist for certain subpopulations. For example, national estimates might be good, while estimates covering smaller geographic areas may be inadequate because of defective geographic coding at the lower (State, County, etc.) level.

### **Specific Error Sources**

As discussed above, errors of undercoverage or overcoverage can be the result of defective frames or of faulty sampling processes. Moreover, the same sources of error can affect both the frame and the selected sample and can lead to either undercoverage or overcoverage. Following are some specific sources of coverage error that are observable and measurable:

**Coding Errors**--Miscoding of industry or Standard Industrial Classification (SIC) coding, geographic coding, size coding, or company structure assignment results in frame errors. Such errors lead either to undercoverage or overcoverage depending on whether the correct units are excluded from the frame or incorrect units included in the frame. Including out-of-scope units (units which should not be included in the sampling frame based on the nature of their business or industrial activity) in the frame results from errors in industry coding and causes overcoverage. By the same token, the exclusion of units of the proper industry results in undercoverage. Similarly, if address, geographic codes, size, or any other attribute is a determinant for the sampling frame, errors in coding will cause overcoverage or undercoverage of the frame.

Two prevalent forms of miscoding are (1) completely unclassified units (especially for SIC) and (2) units which do not have sufficient coding detail for survey purposes. Unclassified units lead to undercoverage since units belonging in the frame cannot be identified. Insufficient coding detail--for example, when four-digit SIC detail is needed and only two- or three-digit detail is available--can lead to either undercoverage or overcoverage for surveys requiring finer levels of industry coding.

Some causes of miscoding are (1) inadequate information on which to base a code; (2) poorly trained coders; and (3) faulty procedures or processes, such as miskeying.

**Errors of Timeliness**--Errors of timeliness result when the frame or sample is not updated to the same reference period as that of the survey. For example, units no longer in business that remain in the frame or sample may lead to overcoverage. Lack of timely updating for new units may lead to undercoverage. For a list frame in which the presence of nonzero payroll is used as an indicator of "activeness," seasonal businesses may be erroneously deleted during their off season. Here again we see the dichotomous nature of coverage error: in surveys which are carried out over time, it is possible to have timely updating of the sampling frame,

but unless the sample, in turn, is updated to reflect these changes, significant coverage error can result. In some survey designs it is impossible to completely eliminate coverage error due to the timing of frame or sample updates. This is especially true for list sample designs. However, use of an area sample to supplement the list sample, such as the Census Bureau uses in its Monthly Retail Trade Survey (MRTS), can theoretically reduce coverage error due to timing to zero.

Structural, organizational, or activity changes not reflected in the frame or sample may occur because of the lack of timeliness in updating. Often SIC changes occur which are not reflected in the frame or sample. Similarly, failure to update for other characteristic changes, such as company reorganizations, acquisitions, and divestments or mergers, results in coverage error.

**Duplication Errors**--Duplicate units on a frame can occur when, for example, a partnership business appears twice, once under each of the partners' identifiers, or when the predecessor and successor establishments both show up as active on the frame, as in the case of a business takeover. This same predecessor/successor situation can affect the sample if one of the units involved is a selected sampling unit. In addition, both a parent firm and its subsidiary could appear as separate sampling units on a frame if the association were not indicated. This would lead to overcoverage if a parent firm and all its subsidiaries are intended to be one sampling unit. Thus, processing or procedural errors can result in duplication error.

Duplication error may also occur when the sampling frame is composed of various lists, which must then be unduplicated. Any error in this process can result in duplicate units being overlooked. This is often a problem where the primary identifiers on the component lists either don't match or are incomplete. Duplication problems also show up in dual frame surveys. For example, in the Census Bureau's Monthly Retail Trade Survey (MRTS), business establishments interviewed by personal enumeration in the area sample must be unduplicated from the list sample frame. When the employer identification (EI) number, which is the primary identifier, is incorrect or missing, the potential for duplication error is particularly great. Here again, while duplicate units cause overcoverage, problems in proper unduplication can also result in a case being incorrectly deleted.

**Deficiencies in administrative record systems, censuses, or surveys on which the frame is based**--Lack of or delays in reporting in the administrative systems, censuses, or surveys can cause coverage error. For example, although firms are asked to submit a separate report form for each of their establishments in the economic censuses of the Census Bureau, some firms invariably provide combined reports on one form. This results in both a deficiency in the frame of multiunit establishments and also in an undercount of the number of business establishments.

**Nonlocatable units**--Sometimes units selected into the sample are not contacted because they cannot be found. In area sample surveys, for example, certain types of businesses, such as service nonemployer establishments, may not be locatable. Noncontact can also occur where street

addresses (for personal interview surveys) or mailing addresses are erroneous or incomplete.

**Interviewer errors**--Errors made by an interviewer in the field can result in the sample being improperly identified. Interviewer "curbstoning" (that is, the interviewer filling out the survey forms without ever properly identifying the establishment or conducting the requisite interviews) and careless canvassing can also lead to an improperly selected sample, loss of population units, or inclusion of erroneous units.

**Processing errors**--Computer programming errors can cause a portion of the selected sample to be omitted from the survey or can result in a deficient frame from which to draw the sample. Units not included due to processing error can also result from poor field procedures or inadequate or incorrect sample maps or materials. Improper identification of the sample at the central sampling facility due to computer or procedural problems can also result in undercoverage. Processing errors (including errors in drawing the sample at the central sampling facility) can lead either to undercoverage or overcoverage.

### **3. CONTROL OF COVERAGE ERROR**

Coverage error can be controlled by many different means. One principle often followed is to identify those areas where coverage error is most serious and assign resources to reduce the error there. Some specific and frequently used techniques which reduce miscoding, lack of timeliness, duplication of units, omission of units, and other errors resulting in incorrect coverage of the sampled population follow:

**Sampling from multiple frames**--Using an area sample to supplement and complete coverage for a list sample is sometimes necessary to obtain complete coverage of the sampled population.

**Integration of multiple lists for frame development**--Integrating and unduplicating several lists to construct a single frame is frequently done since most lists are composites of various sources.

**Conducting special frame improvement surveys**--The Company Organization Survey and SIC classification card mailings for the Census Bureau's Standard Statistical Establishment List (SSEL) are examples of these types of surveys. The economic censuses themselves constitute a frame improvement mechanism for all surveys drawn subsequently from the SSEL.

**Use of two-phase sampling**--This is done in the Census Bureau's business birth sampling program. A first-phase sample is selected based on SIC (including unclassified or insufficiently classified units) and payroll or employment size. A survey is conducted on this sample to produce better coding and to obtain sales data which are used as the measure of size for second-phase sampling.

**Updating for/sampling for births**--Timely updating of the frame and sample for births and deaths.

**Updating for structural changes**--Timely updating of the frame and sample for structural and organization changes of the sampling units.

**Sample validation**--Producing a proof of sample tabulation whereby sample estimates are compared to universe totals for the same characteristic.

**Enlarging the scope of the survey**--Often, in order to capture all of the units relevant to the survey, it is necessary to include possible or marginally possible units. During editing, the out-of-scope units can be dropped.

**Using independent control counts**--These counts are often needed to verify the correctness or completeness of the frame.

**Internal consistency checks for frame content**--This involves performing internal consistency checks on the frame data fields, especially in record identification fields and fields which determine whether the unit is in or out of scope.

**Internal consistency checks for duplicate records**--This procedure involves performing internal consistency checks to identify duplicate records on the frame.

**Include as inscope units with out-of-scope address, geography, industry, size**--The practice of considering as inscope units those which are truly out of scope due to updates or changes in address, geographic, industry, or size code is often used in an effort to represent true inscope units which are not picked up because they are thought to be out of scope.

**Include units closed for the season**--Retaining units closed for a season rather than dropping them and losing their contribution when they become active again is usually necessary to maintain a frame because of the lack of timeliness in reinstating the units.

**Having correct, clear, and manageable sample control and frame maintenance procedures**--All aspects of sample control and frame construction and maintenance must be well thought out and clearly specified.

**Setting up adequate checks on processing**--This is necessary to ensure correct processing of all types: interviewer, clerical, and computer.

**Improving field materials**--Improving field procedures and materials, such as addresses, maps, and other interviewer materials helps to reduce coverage error.

**Interviewer selection and training**--Carefully selecting and training interviewers and coders can have a substantial impact on reducing coverage error. This includes having well-trained supervisors oversee the survey operations.

**Instituting a public relations campaign**--This involves notifying the survey population of the survey or census in advance in an attempt to elicit their participation.

For an example of the procedures which are followed for maintaining frame and sample coverage for a large, ongoing retail trade survey, see Konschnik et al. (1985).

#### **4. MEASUREMENT OF COVERAGE ERROR**

The measurement of coverage error is necessary in all surveys to have some idea of its extent as well as to identify sources most in need of improvement. While the focus of coverage is on the inclusion or exclusion of the proper sampling units in the frame and sample, the measurement of coverage error frequently centers on its effects on the published estimates of the survey. For example, it may be determined that a published estimate for retail sales of establishments in a certain SIC failed to include estimates for a significant number of nonemployer establishments, but that including these nonemployers would only very slightly influence the survey results. The measure of undercoverage would be deemed small despite the number of sampling units excluded.

##### **Indirect Techniques**

Coverage error can often be ascertained by comparing current survey data with results from earlier surveys or from external sources. Coverage error may be indicated if the existing sample shows a significantly higher or lower rate than the comparative data. Such measures as the birth rate, out-of-business rate, out-of-scope rate, unclassified rate, miscoded rate, duplication rate, and sample attrition rate can all be used to identify coverage error.

**Birth rate**--Birth rates may be reviewed, comparing one period to another in order to indirectly measure coverage error.

**Out-of-business rate**--The rate at which frame or sample units go out of business, when compared to other measures or other time periods, provides a useful coverage error measurement.

**Unclassified rate**--A component of coverage error can be estimated by looking at the rate of unclassified units. These when combined with studies of the correct classification of this group provide a measurement of undercoverage.

**Misclassified rate**--A look at this rate and related studies can provide measurements of the extent of coverage error at all levels of survey tabulation.

**Duplication rate**--Determination of the number of repeated or duplicated units in a frame or sample gives useful information on coverage problems.

**Sample attrition rate**--The sample attrition rates, or the rate at which the units in the sample stop reporting over time, provide indications of the extent of coverage error.

## **Direct Techniques**

Direct techniques for measuring coverage error usually entail carefully planned and executed survey procedures designed to provide a reliable estimate of coverage error. The following are examples of these direct techniques:

**Post-enumeration surveys**--Used here, this is synonymous with a post-audit whereby more extensive methods and procedures are used after the conduct of a survey or census in order to identify and determine the effect of coverage errors and other nonsampling errors.

**Matching known population units against frame units**--Checking known population units against the frame provides some indication of the quality of coverage. However, a carefully drawn sample of known units is required before accurate estimates of coverage error can be provided.

**Checking the frame against alternative lists**--While the selected frame may be the best available list for the survey, checks can be made against other lists (either of greater or lesser quality) to measure coverage error.

**Comparing other survey or census data or independent aggregates**--Independent aggregate estimates and tabulations covering the same characteristics for all or a part of the population provide a source of comparison for identifying and measuring coverage error.

**Rechecking interviewers' field work**--Independent rechecks of a sample of interviewers' work are an excellent way of identifying and measuring coverage error.

**Studying components of the frame**--This includes assessing the various classifications of units which make up the list.

## **5. SUMMARY PROFILE**

(See Figures 6a, 6b, 7a, and 7b.) This section presents some general results compiled from the profile of survey practices. Figures 6a and 6b give a summary of control procedures used in descending order of extent of use. Figures 7a and 7b characterize measurements of coverage error taken for these surveys, in descending order of extent of use, for indirect and direct measures.

The results in the figures show that while the majority of these Federal surveys included provisions for controlling coverage error, the measurement of coverage error was less widespread. Moreover, where measurements were taken, only a small percentage was published. Thus, most measurements were for internal use to assess the adequacy of survey estimates.

The most prevalent form of coverage control, used in almost all of the surveys, involved updating the frame for structural changes such as SIC changes, company reorganizations, mergers, etc. Updating of the sample for births was the second most prevalent form of coverage control. Other control techniques reported as being used on more than half the surveys

were: internal consistency checks for duplicate records on the frame; internal consistency checks for frame content; including as inscope units with errors or changes in address, geography, industry, or size, rather than dropping them as out-of-scope; sample validation, i.e., comparison of weighted-up sample units to universe totals; and integration of multiple lists for frame development. Other fairly common control techniques reported were the conducting of special frame improvement surveys and retaining units closed for the season.

Typically, little use was reported of two-phase sampling for improving frames and samples although this method can prove beneficial in reducing the variance of estimates caused by frame problems. Also on the low side in terms of relative use was sampling from multiple frames, such as using both a list and area sample.

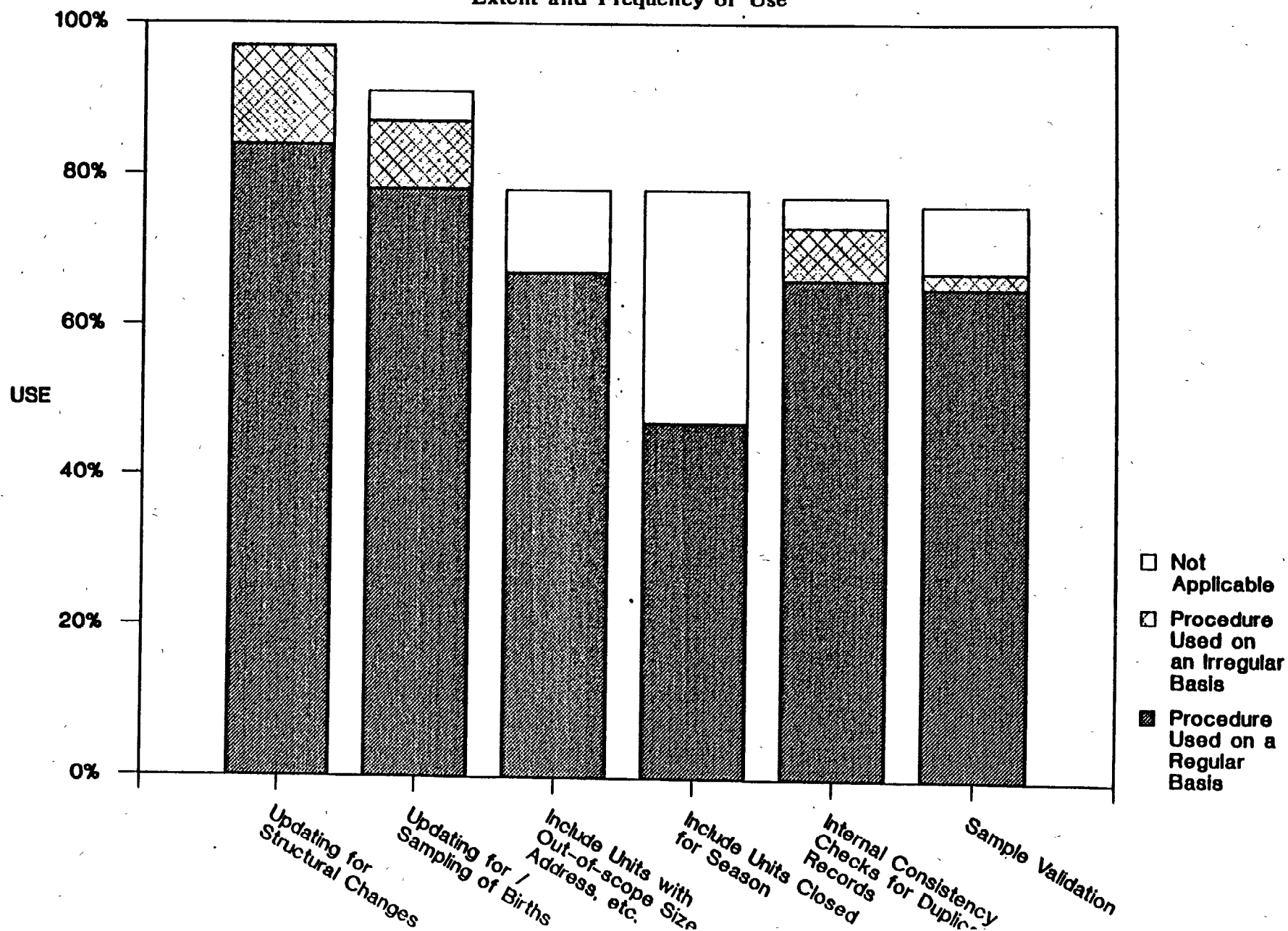
When looking at the measurement of coverage error, out-of-business and out-of-scope rates are most common with around two-thirds of the survey population reported as having these measurements taken, respectively. These measurements also have the highest rate of being published at around 10 percent. A majority of the surveys reported comparing estimates produced in the surveys with estimates based on other independent sources. Measuring the misclassified rate, matching known population units against frame units, unclassified rates, and sample attrition rates were also somewhat common.

Least common were the conducting of post enumeration surveys, presumably because of the resources involved, and rechecks on interviewers' listings, primarily due to the nonapplicability of interviewers' involvement in listing for many of the surveys.

# COVERAGE ERROR CONTROL PROCEDURES

Extent and Frequency of Use

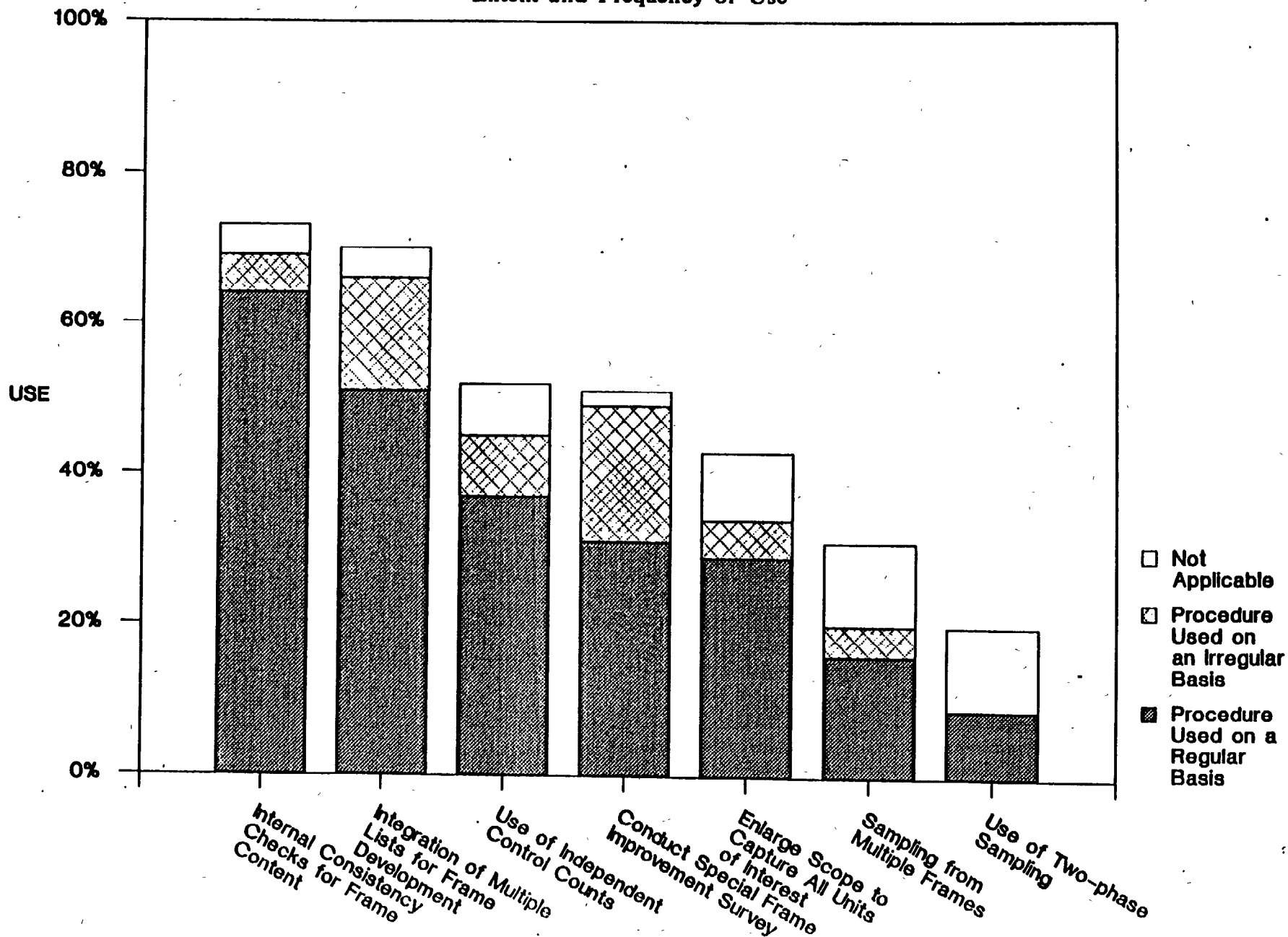
Figure 6a



# COVERAGE ERROR CONTROL PROCEDURES

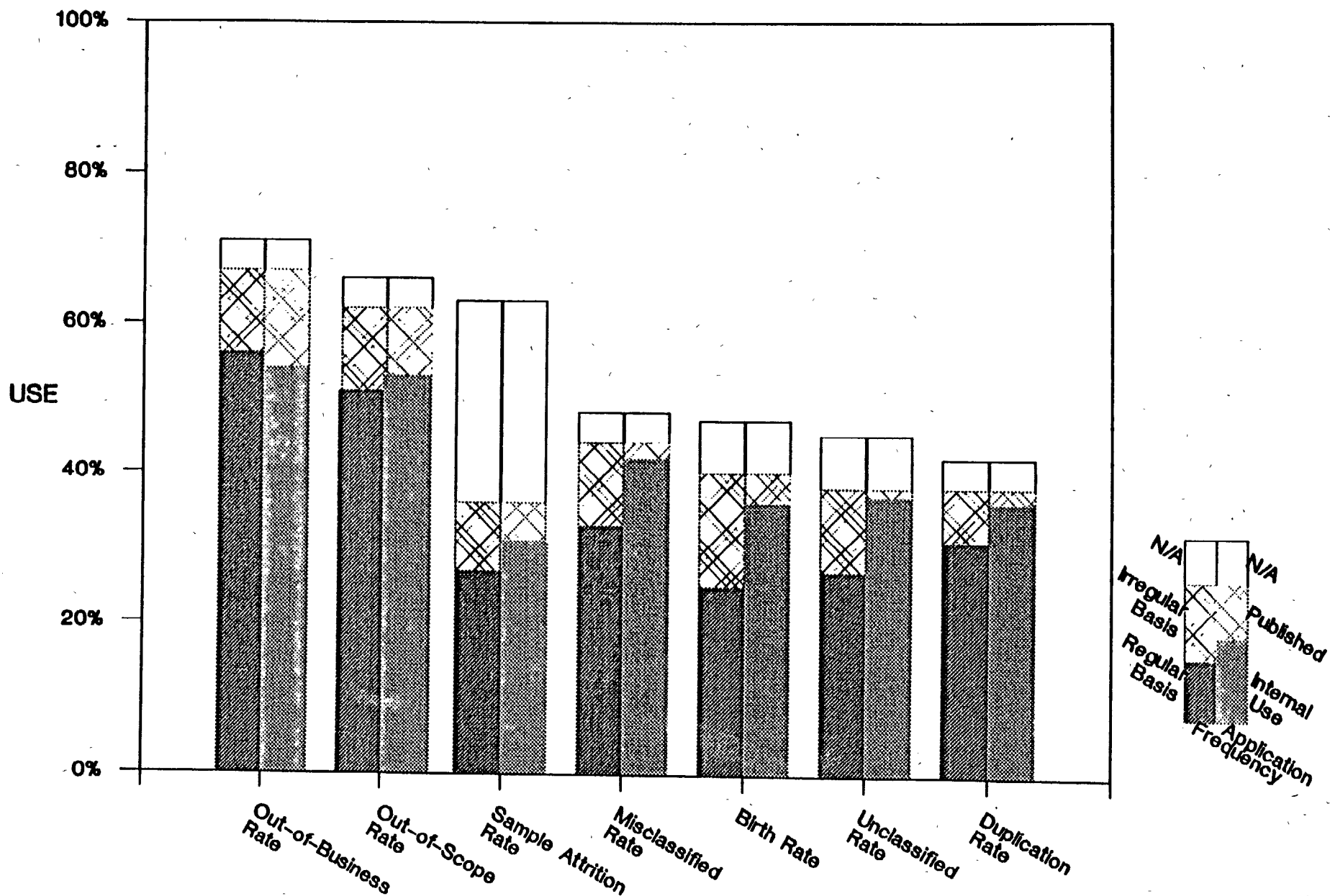
Figure 6b

Extent and Frequency of Use



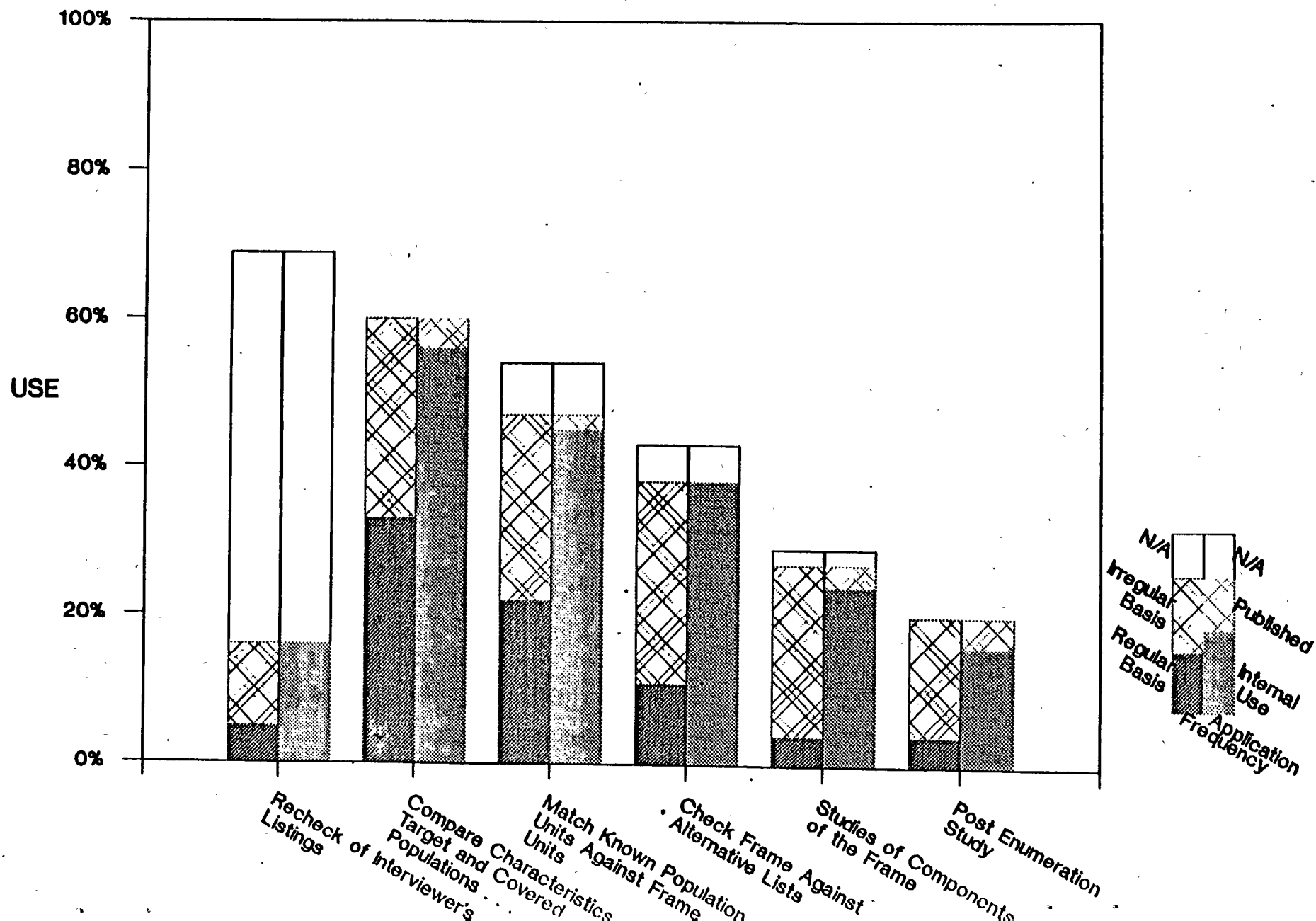
# COVERAGE ERROR MEASUREMENT TECHNIQUES--INDIRECT

Figure 7a



# COVERAGE ERROR MEASUREMENT TECHNIQUES--DIRECT Frequency and Application of Use

Figure 7b



## D. RESPONSE ERROR

### 1. DEFINITION OF RESPONSE ERROR

Response error, which occurs in the data collection phase of a survey, may simply be thought of as the difference between the value collected during the survey and the correct value. Response errors may result from (1) the failure of the respondent to report the correct value (respondent error), (2) the failure of the interviewer to record the value correctly (interviewer error), or (3) the failure of an instrument to measure the value correctly. Although the concept of "correct value" is often simple and well defined, the measurement of the correct value is often difficult and may result in response error. Survey researchers commonly identify response errors as either response deviation or response bias, which is made up of constant bias and variable bias. Constant bias, when it occurs, is a difference between the correct value and the recorded value which is evident over all units in the sample. Variable bias is a change in the difference between correct and reported values for different reporting units. The change in bias may be correlated with the correct value. Response deviation is the component of error associated with differences in the response over repeated measurements of an individual element of the sample. Response deviation is often caused by factors which are unique to the specific interview times, such as, the respondent's attention or the interviewer's actions.

#### Examples

In an agricultural establishment survey, a farmer may report that 160 acres (a quarter of the square mile section which is a common ownership size in the Midwest) are planted in corn when in fact only 154 acres are planted--the remaining 6 acres being roads, streams, irrigation ditches, and the like. This is an example of a respondent error. However, had the enumerator observed the crop growing in the quarter section and recorded 160 acres, the error would be an interviewer error. If interviews at another time or by another interviewer would have resulted in a 154 acre response, the 6 acres would be a response deviation and possibly variable bias. If farmers would always reply 160 acres, the 6 acres are a constant response bias.

Response deviation may occur when several persons who are allowable respondents for the establishment have differing knowledge of the value to be reported. For example, although either spouse is often an allowable respondent for family businesses, one may provide more accurate answers than the other. Thus reported values may depend on which spouse is actually contacted. In establishment surveys, interviews prior to or after completing tax forms may result in response deviations for these data items since the respondent may have more complete financial knowledge after doing taxes.

The simplest example of response bias is when a measurement instrument is miscalibrated. If the error is constant, it would result in a constant response bias. When the error is proportionate to the measurement, there is a variable response bias which is correlated with the correct value.

## 2. SOURCES OF RESPONSE ERROR

The sources of response error in establishment surveys discussed here are grouped into three categories: task error, respondent error, and interviewer error (Bradburn, pp. 289-328 in Rossi, Wright and Anderson, 1983). If an error source is mentioned in only one category, it is done for ease of discussion, and does not imply that sources do not belong in more than one of the categories. Bradburn notes that although "much of the research on response effects has focused on interviewer and respondent characteristics...the characteristics of the task are the major source of response effects and are, in general, much larger than effects due to interviewer or respondent characteristics."

### Task Error

The task is the process of obtaining information. It includes what is measured and how it is measured. The formulation of the task often interacts with the interviewer or respondent to contribute to differences in probing, interviewer or respondent behavior, memory, etc.

A questionnaire of excessive length can cause errors resulting from fatigue or boredom of the respondent or the interviewer. Question sequence can affect the responses when it affects recall or creates confusion.

Questionnaire requirements can also contribute to response error. As mentioned previously, permitting multiple respondents can result in respondents with different knowledge of the desired value and thus contribute to response deviation and/or bias. In situations where multiple respondents are required to complete a questionnaire, the interaction of the group of respondents can cause differences in the reported values.

Records error is a task error which arises from inaccuracy in the records used for responses. Typical causes include inaccurately or incompletely compiled data, the use of inaccurate or out-of-date administrative data, and unavailable or inaccessible records.

### Respondent Error

Respondent error, the failure of the respondent to report the correct value, has many causes. The error may be deliberate or may not be deliberate, as in the case where the respondent does not have adequate knowledge of the establishment data desired. Confusing or lengthy questionnaires or questions requiring extensive data recall or records gathering can also cause respondent error. The burden of reporting is especially worrisome for small establishments that already suffer considerable time loss completing required tax, employment, and other government program forms.

The timing of an interview can also impact respondent error. Interviews soon after the end of a business cycle, tax preparation, or other reporting period may improve recall, while interviews during busy times may result in rushed responses.

Memory problems may occur. Two causes of memory errors are timing and the respondent not considering the requested information to be important. An excessive number of inapplicable questions may cause even the relevant data to suffer. Recall problems include the omission of events or details and telescoping (the inclusion or exclusion of events which are beyond the survey's frame of reference). In establishment surveys in which the respondent is often expected to provide data from records, the problem may be less severe.

The willingness of the respondent to cooperate also affects the accuracy of responses. This may be influenced by the sensitivity of the information, any sense of possible loss of prestige associated with a response, use of the data for taxation or entitlement programs, the respondent's mood, interest in the survey, level of fatigue, available time, sense of burden resulting from repeated visits, and provisions for a tangible or intangible reward for cooperating.

When responses are gathered using a measurement instrument, response errors have been called measurement errors especially in industrial quality control applications. An inaccurate counter, a faulty scale, or poorly calibrated equipment may cause measurement errors. Sometimes weather conditions such as extreme cold, heat, or humidity or physical conditions such as inadequate work areas contribute to measurement errors.

Events that may increase response errors include negative presurvey publicity, adverse legislation or low prices in the establishment's industry, and negative feelings about the survey organization.

#### **Interviewer Error**

Interviewer error, the failure of the interviewer to record responses correctly, commonly results from poor interviewer training or ambiguous guidelines. Deviation from survey procedures is another type of interviewer error. Too heavy a workload may contribute to interviewer error, as does loss of interest in the survey, discomfort with prescribed probing techniques, a negative attitude, fatigue, and inadequate verbal abilities. These factors can cause interviewer error or may result in an interaction with the respondent that promotes respondent error.

The interaction of the respondent with the interviewer or the survey instrument may cause conditioning errors, changes in the response because the respondent perceives a desired answer or realizes that the interview could be shortened, etc.

### **3. CONTROL OF RESPONSE ERROR**

The most common approach to controlling response error is that reflected by O'Muircheartaigh (U.S. Bureau of the Census, p. 209, 1986): "While it is important to assess the overall quality of the data in a survey, it is frequently a greater concern to identify particular problem areas. Some variables will be more susceptible to unreliability in reporting than others, and some classes of respondents will be less consistent than others in their responses. It would be useful to identify these variables

and these types of respondents and to examine the reasons for the lower quality of data they provide.

"Having identified problem areas the next stage should be to change the survey procedures to take the problems into account and if possible to overcome them. This might involve changes in the definitions of, and questions for, the constructs being measured and/or changes in the field work strategy and execution. Such changes are more appropriate in the context of a continuing survey (or of a program of related surveys) than in a single ad hoc survey. In a continuing survey it is possible to monitor the impact of the changes by continuing to evaluate the data after the changes have been introduced."

Some techniques for controlling the previously mentioned sources of response error in establishment surveys follow:

#### **Task Error**

Some basic methods used to control questionnaire misspecification include studying establishment recordkeeping practices prior to designing the survey forms, attempting to understand how respondents interpret the questions and answer them, and using questionnaire pretests. Working Paper 10 (Statistical Policy Office, 1983) provides detail about controlling questionnaire misspecification. Techniques used include: individual and group interviews, interview observations, formal testing, and post survey evaluation.

Studies to check records and to eliminate nonmeasurable data items from the survey or to improve collection methods are useful ways to control records error in establishment surveys.

#### **Respondent Error**

A simple method of controlling respondent error in establishment surveys is to check responses against administrative data when they exist. An analyst familiar with the industry may be able to spot responses which are uncharacteristic of establishments in the industry with similar administrative data.

Where respondents must provide data in repetitive contacts, personal contact with the respondents whose data often contain problems may help improve responses. Finally, a computer edit which utilizes all reasonable relationships within the record is essential, as are effective followup procedures.

Recently, techniques from cognitive psychology have been used to study sources of respondent error. (See Loftus and DeMaio, et al. in U.S. Bureau of the Census (1986).)

#### **Interviewer Error**

The control of interviewer error starts with detailed and understandable training and procedural guidelines for the interviewers. The management aspects of a survey--recruitment, training, and supervision of the

enumerators--must receive proper attention. Testing and well defined, relevant selection criteria during interviewer recruitment can control interviewer error. Supervision practices will vary with the survey conditions such as telephone vs. personal interviews, number of interviewers supervised, etc. Development of good supervisory practices is essential because the supervisors are often the first level at which problems are recognized or corrected. Supervisors can help interviewers understand their job better, provide additional training, and assure that workload does not impact the quality of the work. Field editing may be useful, or when using telephone interviews, on-line monitoring is useful. A reinterview of a sample of the interviewer's work is also a commonly accepted practice.

#### 4. MEASUREMENT OF RESPONSE ERROR

Since the sources of response error are extremely diverse, the techniques for measuring it are also diverse. Measurement studies have been conducted to:

- (1) estimate the precision of survey results,
- (2) identify specific survey problems,
- (3) identify improvements in the survey methodology, and
- (4) monitor the impact of changes to the survey methodology.

The following is a generalization of some of the measurement approaches taken in studies of response error.

The measurement of response errors requires that they be represented by a mathematical model. A number of alternative models have been proposed, often to accommodate special situations. Most sampling textbooks provide an example of an error model and further references. To illustrate, a general response error model (Cochran 1977) is

$$y_{ij} = x_i + e_{ij} \\ = x_i + b + b_i + d_{ij}$$

where  $y_{ij}$  is the value obtained from the  $i$ th element in the  $j$ th repetition,

$x_i$  is the correct value,

$e_{ij}$  is the error of measurement,

$b$  is the constant bias term of  $e_{ij}$ , if any,

$b_i$  is the variable component of bias which may be correlated with  $x_i$ , and

$d_{ij}$  is the fluctuating component of error from repetition  $j$  which follows some frequency distribution.

The variations in the response error models which have been developed depend upon the survey itself, the error sources assumed to be a problem in the survey, and the assumptions made about  $e_{ij}$ . Survey factors which must be considered by the model formulation include (1) the existence of, or ability to obtain, "correct" values for units in the survey, (2) the complexity of estimation from the sample design, (3) the ability to make

remeasurements under reasonably fixed conditions, one of the most difficult conditions to achieve, (4) the ability to randomize work assignments, and (5) budget constraints for these costly measurement studies.

The predominant method of measuring response error involves formulating a response error model, postulating that the survey is repeatable under some fixed set of identical conditions, and measuring the components of variability (response variance) among the repetitions. Interpenetration and reenumeration (or a combination of the two) are commonly used to measure the response variance. Fellegi (1964) presents a framework for the joint application of these techniques while Cochran (1977), Wright (1983), Zarkovich (1966), and the U.S. Bureau of the Census (1985) provide numerous references to approaches taken in different circumstances. A discussion of reinterview methods, sometimes called response analysis surveys, can be found in Working Paper No. 10 (Statistical Policy Office, 1983).

Measurement techniques can also be used as a control method. This approach involves controlling the survey estimates by adjusting the survey estimate to counteract the bias. Zarkovich (1966) recommends double-sampling approaches which estimate response bias. Basically, this approach consists of selecting a subsample of the original sample, collecting "correct" values for these responses, and forming a difference estimator using the original responses. A limitation is that the "correct value" which is necessary for the approach is often not obtainable. Examples of double sampling can be found in Tenebein (1970) and Ostry and Sunter (1970).

Measurement techniques include both indirect and direct measurement techniques.

#### Indirect Techniques

Indirect measurement of response error involves examining the information related to response error. This includes the usual survey practice of computing edit failure rates and interviewer error rates. This type of information does not measure the response error, but does provide a reasonable idea of the magnitude of the error. Feedback sessions with respondents and/or interviewers may also help find sources of response error. Questionnaire pretests and cognitive studies, which among other things can help determine whether different word meanings are assumed by different respondents or how recall methods affect response, also provide clues concerning the magnitude of response errors.

#### Direct Techniques

Direct measurement of response error requires a designed study. The study may be as simple as a records check or may be a detailed content or reinterview study that attempts to control causes of error. Interviewer and respondent variation studies often assume that an identical set of survey conditions have occurred during repeated or randomized assignments of data collection by the interviewer or in repeated inquiries from the respondent. Under such conditions the contribution to error from interviewers or respondents can be measured.

## 5. SUMMARY PROFILE

Figures 8a, 8b, 9a, and 9b illustrate control procedures used and measurements produced to evaluate response error based on the profile of survey practices.

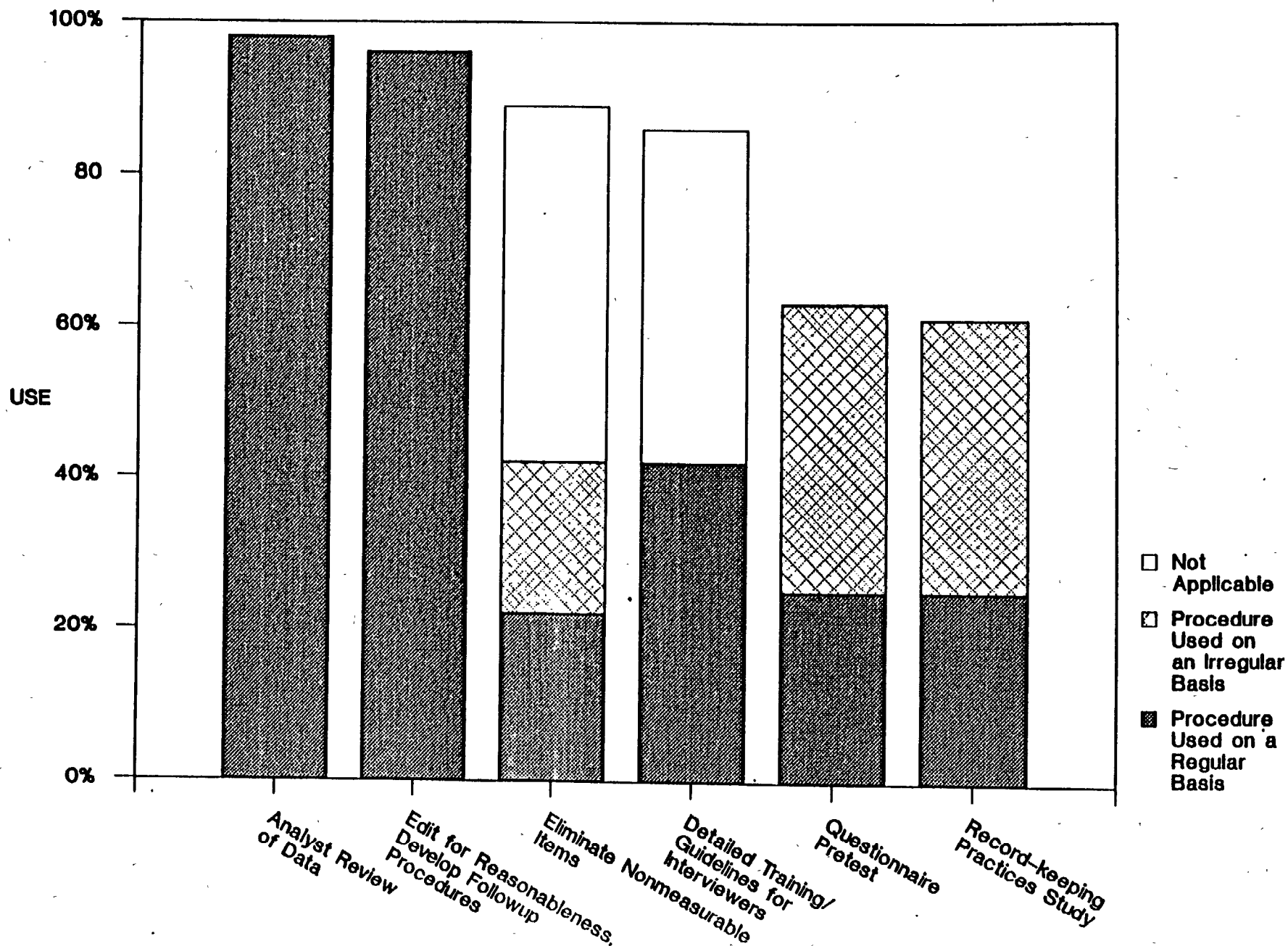
As might be expected, virtually all surveys reviewed by this report indicated that an analyst review and data edit were used to control response errors. Unfortunately, inspection of a single response can usually detect only the most extreme response errors. Reinterview studies were uncommon, but about half of the surveys conducted administrative data and/or records checks. The surveys which address response error (about half of those for which this information was collected) concentrate efforts in the planning and execution stages of the survey by using record-keeping practice studies, questionnaire pretests, detailed training for interviewers, and personal visits. Cognitive studies and CATI on-line monitoring, which have been much discussed recently by survey researchers, are a part of only a small fraction of the surveys.

About three fourths of the surveys produce edit failure rates to indirectly measure response error. Yet, fewer than half of the surveys provided applicable detail about cause of the response error such as interviewer error, records checks, or response variance. It was interesting to note that questionnaire pretests and cognitive studies were indicated as producing response error measurements at only half the rate for which they were reported as a control procedure.

# RESPONSE ERROR CONTROL PROCEDURES

Figure 8a

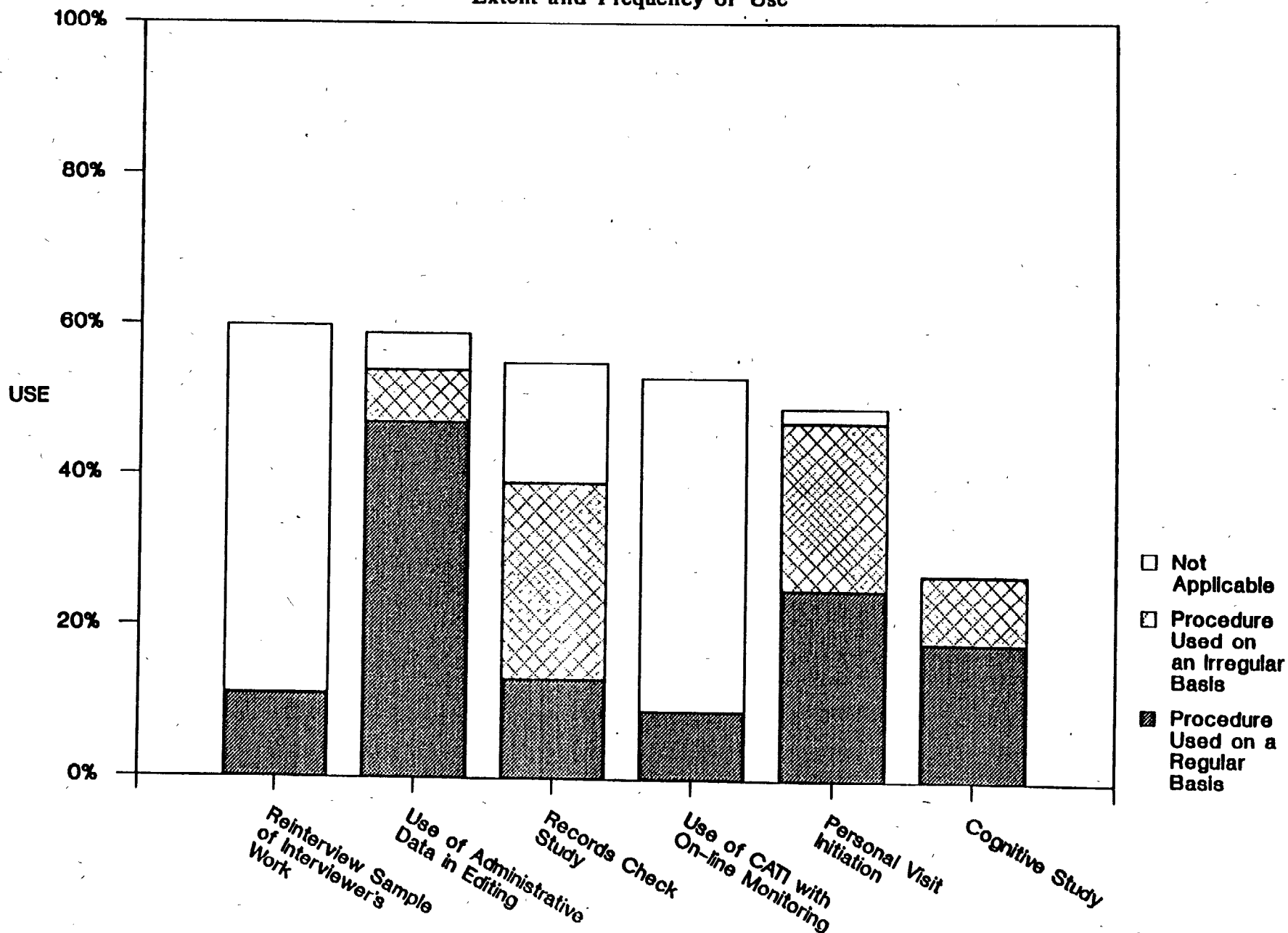
Extent and Frequency of Use



# RESPONSE ERROR CONTROL PROCEDURES

Figure 8b

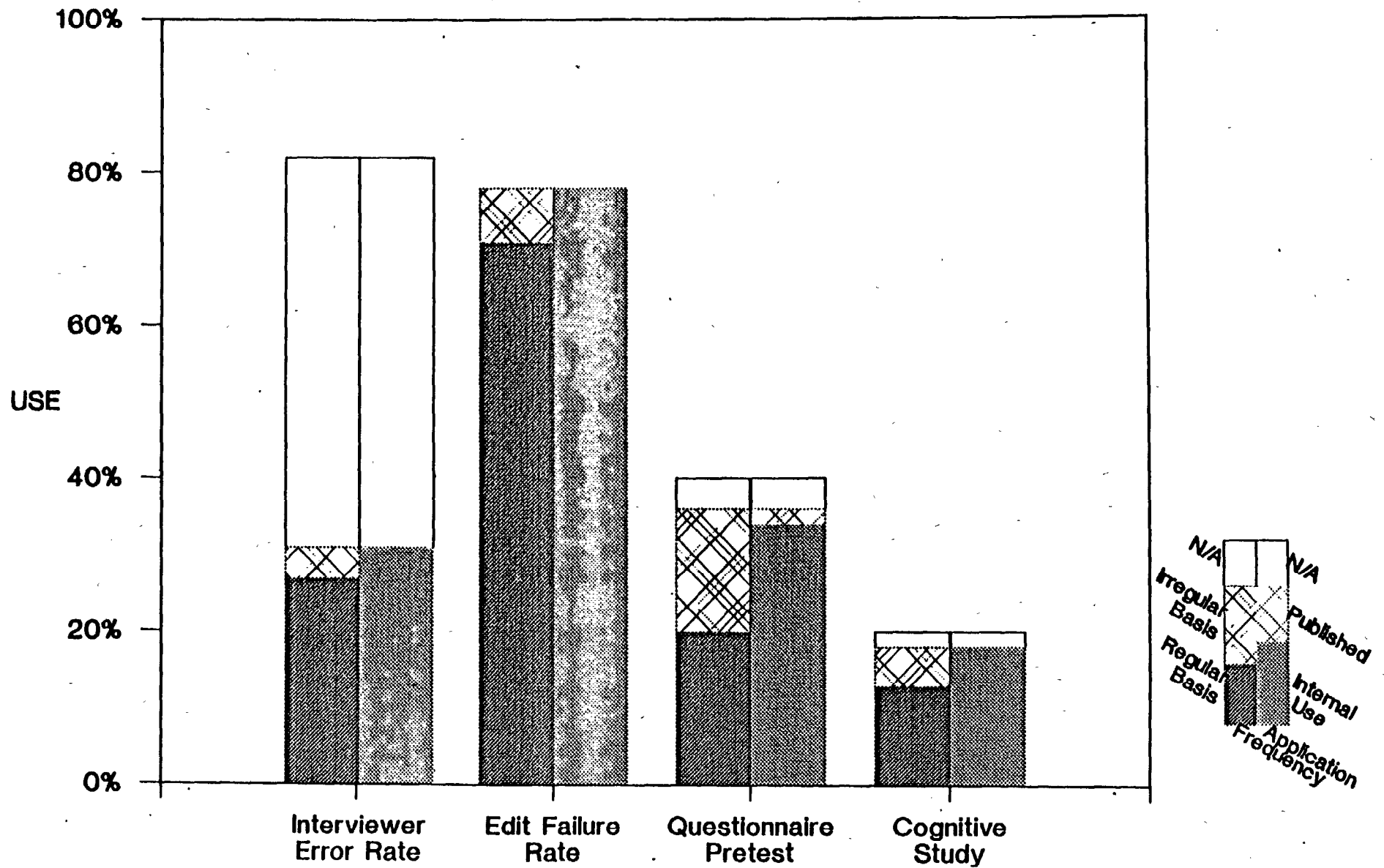
Extent and Frequency of Use



# RESPONSE ERROR MEASUREMENT TECHNIQUES--INDIRECT

Frequency and Application of Use

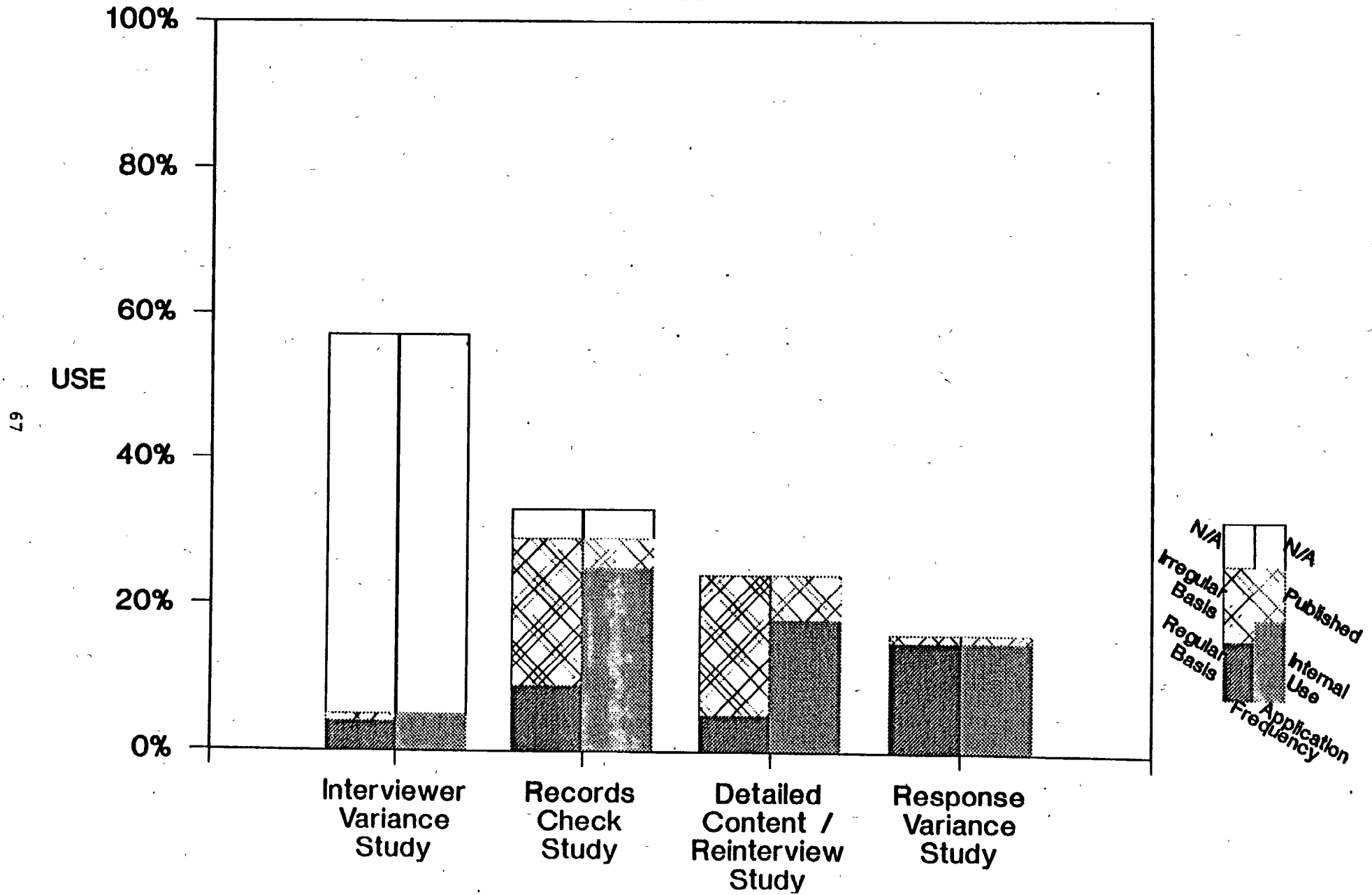
Figure 9a



# RESPONSE ERROR MEASUREMENT TECHNIQUES--DIRECT

Frequency and Application of Use

Figure 9b



## **E. NONRESPONSE ERROR**

### **1. DEFINITION OF NONRESPONSE ERROR**

Nonresponse error results from a failure to collect complete information on all units in the selected sample. Nonresponse produces error in survey estimates in two ways. First, the decrease in sample size or in the amount of information collected in response to a particular question results in larger standard errors. Second, and perhaps more important, a bias is introduced to the extent that nonrespondents differ from respondents within a selected sample.

In Sections 2 through 5, respectively, we will look at some of the sources of nonresponse error in establishment surveys, examine techniques for controlling the error, discuss methods for measuring the extent of the nonresponse problem, and present a summary profile of current nonresponse error techniques at government agencies.

An excellent reference on survey nonresponse error is Madow et al. (1983), especially Volume 1, which presents a comprehensive discussion on the subject.

### **2. SOURCES OF NONRESPONSE ERROR**

There are three primary sources of nonresponse and they can be represented as a hierarchy. First, a sampled company may not be contacted, in which case the establishment does not have an opportunity to respond. This is referred to as a noncontact. Second, a sampled unit that is contacted may fail to respond. This represents unit nonresponse. Third, the unit may respond to the questionnaire incompletely. This level is referred to as item nonresponse.

#### **Noncontacts**

When an attempted contact of a selected survey unit results in a failure to contact or when no contact is attempted, the nonresponse is classified as noncontact. One failure to contact that could occur in establishment surveys results from seasonal closings (for example, in the vacation and leisure industry, with seashore resorts closing during the winter and ski resorts and ski equipment shops closing for the summer--and the food processing industry, which is affected both by seasonality and disturbances in the weather).

An attempted contact may also fail because of a temporary closing due to a strike or work stoppage, a possible event in industries with strong and radical labor unions.

Attempted contacts may not succeed due to a failure to locate the company. The firm may have moved or changed telephone number, or an incorrect address may have been inserted on the universe file. In the case of mail surveys, the survey form might be sent to the wrong location, the form misplaced prior to mailing, or lost during the mailing process.

Nonattempted contacts may result from negligence or sabotage on the part of the interviewer or in the mailing operation. Also, there may not be enough time in the collection period to reach all sampled units. The end result is that the sampled company is never contacted in the first place.

### **Unit Nonresponse**

Once the sampled company is contacted, lack of any response to the questionnaire is classified as unit nonresponse. It is simply the failure of a contacted company to respond. Here again, certain sources of unit nonresponse are common to establishment surveys. For example, the survey form may never reach the appropriate division or contact person. This is most likely for large conglomerates with many divisions in diverse locations. The headquarters of a large corporation might be in a different city, or even a different State, than the production divisions.

Another source of unit nonresponse is when the sampled company is participating in too many surveys. This is especially true among the largest establishments, which because of their size may be included in every survey of their industry. Smaller companies, although not as likely to be involved in numerous surveys, may also have trouble finding the time to respond due to limited staff and resources.

Excessive costs of retrieving data is another reason for unit nonresponse among establishments. For example, a survey might ask for a particular disaggregation from company files that would require creating a new program to assemble the data. Another problem is that a company may have complex file structures that do not lend themselves to easy retrieval of the data in the form that the survey requests.

In other cases, the data requested may not be relevant, or the contact person decides it is not relevant to the company and tosses out the form. Also, unit nonresponse results from units being unwilling to cooperate; some companies might have a blanket policy of not responding to voluntary surveys or confidentiality of the data could be an issue.

### **Item Nonresponse**

Item nonresponse is the failure of a responding unit to answer a particular question. As with unit nonresponse, excessive costs are a primary cause of item nonresponse. Respondents might answer those questions that can be answered easily and skip over those requiring expensive data retrieval and manipulation.

Item nonresponse may also result from technical difficulties. For example, some data may not be available during the survey period due to the ongoing development of a computer system to retrieve and assemble the information. Other times data may be unavailable due to systems processing problems at the time of the survey. Of course, if the problems are widespread, the result may be unit nonresponse.

Sometimes item nonresponse may reflect deficiencies in the questionnaire. Surveys that request too much data are apt to yield many partial returns. Questionnaires that are complicated, look cluttered, or have ambiguous questions or unclear instructions have increased probability of item nonresponse (or even unit nonresponse). Sensitive questions or queries in areas the company regards as confidential may also be omitted.

Another source of item nonresponse may be the fault of the interviewer who does not follow the instructions provided or may either purposely (for example, because of time constraints) or accidentally omit questions.

### **3. CONTROL OF NONRESPONSE ERROR**

#### **Noncontact**

First, to reduce noncontact of sampling units, controls can be instituted to ensure a strong effort to produce a successful first wave of contact and persistent followup procedures in the event of initial failure. In the case of mail surveys, mailing lists should be carefully checked to obtain accurate addresses. Annual or quinquennial benchmark surveys may require extensive research to update and verify mailing lists. Establishing process and quality control procedures on the mailing operation can further ensure that all survey forms are mailed and then received by sample units. Quality control procedures on the mailing operation are used in about 91 percent of agency mail surveys.

For interview surveys, interviewers who are convinced of the importance of the data collection effort will make an extra effort to reach all sampling units. Intensive followup of establishments identified as critical to the success of the survey is widely done at agencies both for interviewing and in the mailing process.

#### **Unit Nonresponse**

The distributions of companies in many establishment surveys are highly skewed. For example, the BLS distribution of establishments by number of employees given in Chapter III showed that about 2 percent of all establishments contain around half of all employees. Another example is the population of finance companies where about 10 percent of all companies report over 90 percent of total lending.

Given these settings, it is clear that large companies in the sample are critical to the success of the survey. Thus the followup of large companies who are not responding is very important. Followup techniques may take the form of reminder cards, periodic telephone calls, or re-interviews. Over three-fourths of agencies' surveys employ intensive followup of critical units. And nearly all surveys queried use some type of unit nonresponse followup procedure.

Giving advance notification to a selected company can reduce nonresponse rates. This is more important in establishment than in household surveys due to the bureaucratic organization of most companies. Sending a letter informing companies of their selection for the survey, a statement of its purpose, and a cordial request for their cooperation may be helpful. A

personal visit or telephone call by a member of the survey staff to selected establishments may also be effective. Results from agency experience are indicative that a strong effort is put forth to introduce the survey to the selected sampling units.

Another good front-end technique for promoting cooperation is to offer the company a copy of the statistical release or published survey results if they participate. This appears to be a standard technique at government agencies as the survey tabulation shows.

The use of special reporting arrangements may encourage large companies to respond. Large companies that are vital to the survey because of their large holdings of key survey variables may appreciate special treatment. For example, suppose a survey is conducted out of Washington, D.C., but the data collection is done through district reporting centers. It may be beneficial to offer large companies direct communication with headquarters or central office clearance. This not only allows them more time to prepare the data, but eliminates an intermediate step in the event that problems occur with the reported data.

For surveys that collect detailed information, large firms may have thousands of observations, whereas small firms may have only a handful. Special arrangements to encourage the cooperation of the large firms may include allowing them to submit data on magnetic tape, floppy discs, or according to a specially arranged format.

Special care and treatment may also be necessary to produce a good response from the smallest sampling units. Unless the survey is short and simple, small companies that respond may face a disproportionate cost due to their limited resources. Responding to a complex survey, whether done manually by internal staff or by hiring outside programmers (possibly requiring the purchase of more sophisticated data processing equipment), may be a significant financial burden.

After reporting strong use of controls at the front-end of the survey operation, government agencies continue to pursue small nonresponse rates by using special reporting arrangements in about three-quarters of the operations.

Another control technique for increasing the response rate among small establishments is sample rotation. A company participates in the survey panel for an agreed-upon length of time and is then replaced by another company having similar characteristics. Where applicable, this technique is used in around one-third of agency programs.

Survey designs where adherence to a strict probability selection is important may need to be changed from time to time because of a shifting population, perhaps due to growth or geographical relocations. This requires a redesign with assignment of new probabilities of selection to population units. Maximizing the overlap across survey designs may be desirable in order to provide stable, comparable data series. Additionally, sizable investments may have been made by both respondents and agency in order to collect the data. There are a number of techniques available, including the use of certainty selection and the use of

conditional probabilities based on the previous design. Two references giving techniques for changing from an initial set of probabilities to a new set are Keyfitz (1951) and Kish (1967).

### **Item Nonresponse**

Once the selected company commits to participation, the final step is to ensure that it answers all survey items. An important part in reducing item nonresponse is played by the a priori knowledge of the data storage structures of establishments in the sampling frame. Acquiring this knowledge may require a pilot test or presurvey questionnaire. This could ask for such things as how the requested data are stored, if the response will be manual or computerized, if data can be disaggregated, or if the data can be retrieved and assembled in the form desired. Then using the results of the pilot test, the survey questionnaire can be tailored to fit the recordkeeping practices and abnormalities of the surveyed population. The agency survey showed that nearly half the applicable programs employed a data-keeping pilot test.

Item nonresponse followup appeared to be in widespread use at government agencies. An important factor here is the training of interviewers and data editing clerks in the importance and use of the survey data. Additional patience may be required in collecting items from establishments due to the many tiers of personnel. A circuitous path may be encountered before a correct contact is made. The use of nonresponse measures can also be an aid in followup procedures. Item nonresponse and item coverage rates can flag key items that need callbacks.

The design of the questionnaire is another factor in controlling item nonresponse. Since poorly organized survey forms, poorly illustrated questionnaire skip patterns, and excessively long questionnaires are known to increase item nonresponse, a clear unambiguous survey form that can be completed in a reasonable amount of time is beneficial.

## **4. MEASUREMENT OF NONRESPONSE ERROR**

Various measures of nonresponse error can be assembled at the data processing stage of a survey. There are both direct and indirect measures and indicators that can be used to assess the effect of nonresponse on the survey. Direct measures produce estimates of the bias in survey estimates due to nonresponse. Indirect measures do not provide an actual estimate of the bias, but do give some indication of the possible existence of nonresponse bias and its seriousness.

### **Indirect Techniques**

The unit response rate is frequently used as an indirect measure of nonresponse. Easy to compute, it is the ratio of the number of responding eligible units to the number of eligible units in the sample. The unit nonresponse rate is of course the complement of the unit response rate. During the data processing stage of the survey, this measure provides a useful warning sign of the extent of the nonresponse problem. Later, when survey estimates are available, these rates provide indicators of nonresponse bias. The agency practices survey showed a strong use of

unit response rates, over three-fourths. However, it is interesting to note that only about one-fourth of these surveys actually publish the rates with survey estimates.

In establishment surveys, a better analysis of the nonresponse problem can be obtained by tabulating unit response rates by size of institution. For example, a 95 percent overall response rate is not as good as it appears if only one of the 10 largest companies responded. The tabulation of unit response rates by interviewer or geographical area may also identify problems with the data collection effort.

Weighted unit response rates, a refinement of unit response rates, are particularly valuable in establishment surveys. The frequency distributions of economic variables such as income or expenditures are highly skewed for many establishment populations. If the weighting variable is income, then a 50 percent unit response rate could translate to a 90 percent weighted unit response rate, for example. Here the agency survey showed that around half of programs computed weighted response rates but only about a sixth of these publish the rates.

Item response rates are indirect measures of nonresponse on a micro level. They are calculated as the number of eligible units responding to an item divided by the number of eligible responding units. Just over half of agency programs report use of these rates with roughly a sixth actually published. These rates provide an early indication of nonresponse and may be helpful if shown by size of industry, interviewer, geographic area, or some other stratified variable.

The item coverage rate may be more useful than the item response rate in establishment surveys. Defined as the ratio of the total of a significant variable (for example, income, acreage, total deposits) for eligible responding units to the total for all eligible units in the sample, it is a meaningful measure of nonresponse in establishment surveys where a relatively small number of firms have a disproportionately large share of the market. The table of agency responses shows less use of this measure than unit and item response rates but a relatively higher fraction of publication.

The refusal rate, measured as the number of eligible units that refuse to participate divided by the number of eligible sample units, provides indirect information about the willingness to respond among the population of companies. This could say something about the difficulty of the questionnaire, the unit contact and reception process, or the ability of the interviewer. Improved information results from this rate being tabulated, whether by interviewer, collection district, State, or other entity. About half of agency surveys compute refusal rates.

Knowing the reason for either unit or item nonresponse is helpful toward obtaining future reductions in the nonresponse rates. This understanding can be incorporated in a redesign tailored to correct survey response difficulties. The agency surveys results show that about a quarter of all applicable programs keep some type of data base on the reason for the nonresponse.

Nonresponse adjustment is typically carried out using data obtained from one interview period. In the case of a panel survey, data collected across interview periods may be used to evaluate the nonresponse adjustment procedure. This longitudinal data may also be useful in developing models for or refining the nonresponse adjustment procedure. About a fifth of agency surveys report using data across survey periods for nonresponse adjustment.

### **Direct Techniques**

A direct measure of nonresponse bias is obtained by collecting some of the survey data or covariate data for nonrespondents from another source, such as from a census or from administrative records. Comparisons with respondent census data by various subgroups yield differences which make possible the construction of correction factors to adjust for nonresponse. The characteristics of most establishment populations make the formation of subgroups important in determining differences between respondents and nonrespondents, for example, large companies versus small ones. About a quarter of agencies programs make direct nonresponse adjustments based on administrative data for nonrespondents.

Another way of deriving a direct measure of nonresponse bias is to draw a sample of nonrespondents and conduct an intensive followup to collect the data. Estimates of the nonresponse population are constructed from this sample and compared to those based on the respondent sample. Differences between the two populations are a measure of the nonresponse bias. This technique is sparsely used in agency programs.

### **5. SUMMARY PROFILE**

(See Figures 10a, 10b, and 11.) Respondents to the government-wide questionnaire supplied data on control procedures used to contain nonresponse error. Error source categories are the three given in the text: noncontact, unit nonresponse, and item nonresponse.

In surveys where a mailing operation is involved, the use of controlled procedures to ensure the accuracy of the mailing operation is nearly unanimous.

Generally the data show a strong effort in the preliminary work of encouraging survey participation. Especially notable are advance notification efforts and special reporting arrangements for critical establishments. The more costly programs of personal visit initiation and pilot testing for recordkeeping practices show somewhat less use.

Frequent use is made of followup procedures for unit and item nonresponse and in contacting establishments deemed critical to the success of the survey.

Data on nonresponse measurements are separated by type into indirect and direct measures. Among the indirect measures perhaps the most striking results are the differences between the use and the publication of the measures. We find there is fairly strong use of certain nonresponse rates but frequently these nonresponse rates are not published. Moreover,

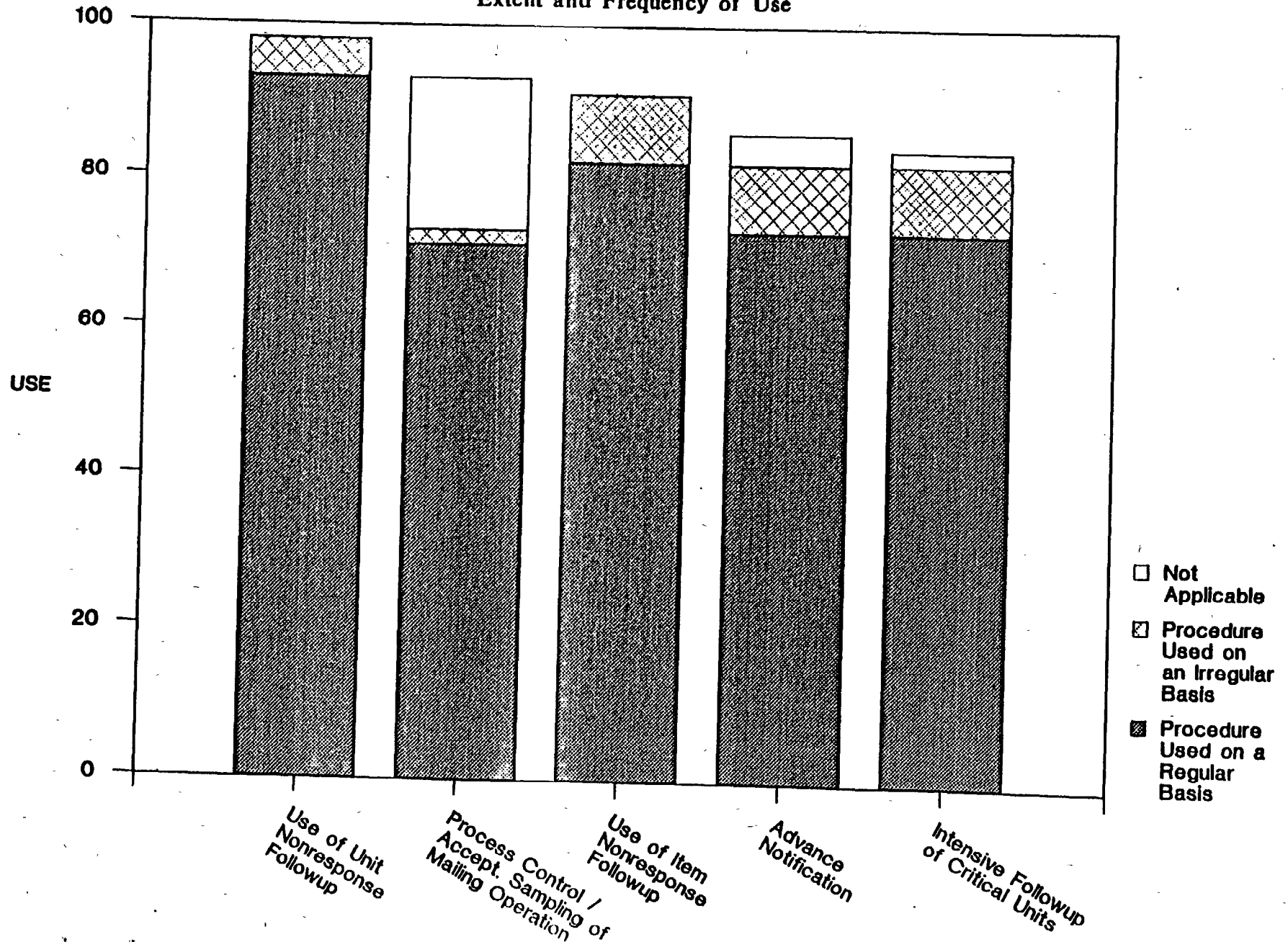
there does not appear to be a strong effort to record and document the reason for the nonresponse.

The direct measures of linking to administrative data and a followup sample of nonrespondents are sparsely in use. Perhaps the complexity and the costs associated with these measurement types are the primary reasons for this.

# NONRESPONSE ERROR CONTROL PROCEDURES

Figure 10a

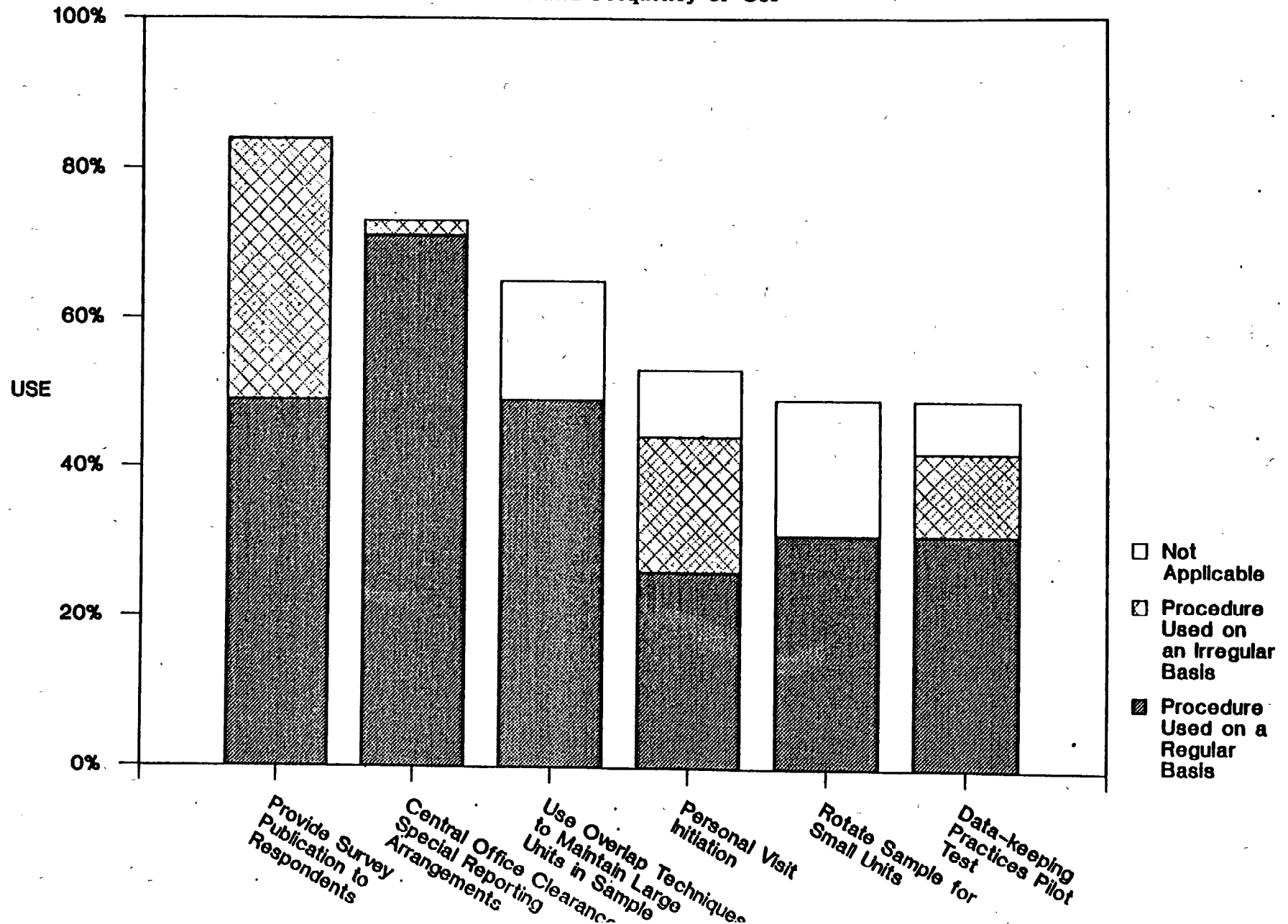
Extent and Frequency of Use



# NONRESPONSE ERROR CONTROL PROCEDURES

Figure 10b

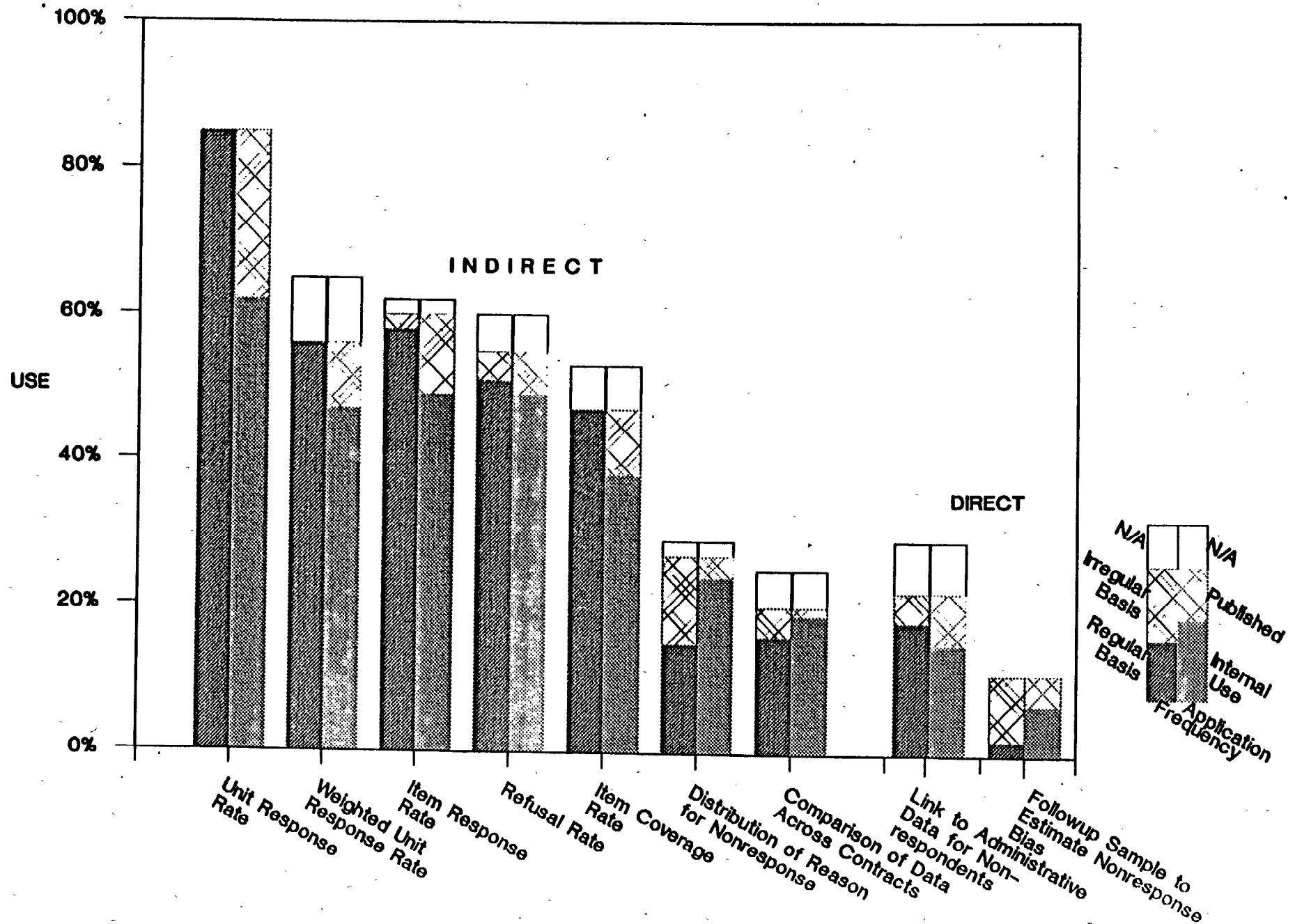
Extent and Frequency of Use



# NONRESPONSE ERROR MEASUREMENT TECHNIQUES

Frequency and Application of Use

Figure 11



## **P. PROCESSING ERROR**

### **1. DEFINITION OF PROCESSING ERROR**

Processing error is the error in final survey results arising from the faulty implementation of correctly planned survey methods. As discussed here, processing errors encompass all post-collection operations, as well as the printing of questionnaires. Most processing errors occur in data for individual units, although errors can also be introduced in tabulations and estimates.

### **2. SOURCES OF PROCESSING ERROR**

Instead of compiling a lengthy listing of processing errors, we will categorize the major sources of such errors--namely, the preparation of the questionnaires, the data collection process, the clerical handling of the forms, and the processing of the data by clerks, analysts, and computers. Basically, these categories cover any processing problems from the printing of the questionnaires to the publication of survey results. Some processing errors affect the quality of the survey results directly (keying errors, for example), while others have indirect effects (poor printing on mailing labels, for example, which could lead to increased nonresponse). Generally it is difficult to completely separate the effects of processing errors from the effects of nonresponse, response errors, and coverage problems. Moreover, the categories of processing errors used here are not intended to be mutually exclusive since interactions between processing activities can cause more errors. For convenience in discussing processing errors it is assumed that the sample design is correct and that both the questions being asked of respondents and their responses are correct.

#### **Questionnaires**

Even after a draft questionnaire has been carefully field tested, errors can creep in during the final preparation and printing. For example, arrows indicating skip patterns or boxes for checking the appropriate response may be dropped, typographical errors may occur, or question and answer boxes may be poorly arranged, any of which can make it difficult for the respondent or interviewer to complete the form. Printing errors such as pale or smeared type may also decrease the response rate. These types of problems occur most often when a large number of similar forms must be prepared and printed at the same time, such as the economic censuses, for which a basic questionnaire is tailored to each of several hundred SIC categories. A few people must proofread and review a large number of questionnaires in a short time, leading to reviewer fatigue and errors. Any of the problems mentioned here can result in erroneous or missing data.

#### **Data Collection Process**

Many processing errors can occur during the actual collection of data from respondents whether the data are collected by mail, telephone, or personal visit. For example, the wrong type of form may be mailed to a respondent, or a telephone interviewer may not follow the questions on

the questionnaire correctly. Even when data collection procedures are carefully spelled out, the following types of errors can occur: For mail surveys the form may be sent to the wrong location or the form may be sent to an inappropriate person within the company. For telephone or personal visit surveys the wrong unit may be called or visited, data may be collected from an inappropriate respondent, the interviewer may "lead" the respondent to a particular answer, the interviewer may "second guess" or assume answers, a question may be skipped, or the interviewer may probe in an inappropriate manner or may fail to probe.

The special difficulty associated with data collection errors is that the results are usually indistinguishable from nonresponse and response errors. The agency sponsoring a survey will not be able to distinguish a nonrespondent who chose not to respond from a nonrespondent who didn't receive the form because it was sent to the wrong location. Similarly, the survey taker can't separate true response error (that is, the respondent providing erroneous data) from erroneous data caused by an interviewer asking the wrong question. Because of this, the processing errors that occur during contacts with respondents are usually treated as though they were nonresponse or response errors.

#### **Clerical Handling of Forms**

Many opportunities for mistakes that can affect the quality of survey data arise in the handling of the questionnaire forms. Before mailing, questionnaires may get sorted by company, SIC, geography, and zip code, and forms and instructions must be folded and stuffed into envelopes. Errors in these activities lead to nonresponse problems (which were discussed in detail in a previous section). After mail returns, envelopes are opened, forms are checked in (clerically, by keying, or by bar code reading) and sorted. During all the shuffling, forms or instructions can be left out of a mailing piece, forms or parts of forms can get lost or damaged, forms can be checked in under the wrong identification, may not get checked in, or may get checked in more than once. These mistakes lead to nonresponse, duplicate response (from unnecessary nonresponse followup), lost data, and data stored under the wrong unit identifier.

#### **Data Processing by Analysts and Clerks**

Clerical and professional staffs are responsible for many activities that provide opportunities for mistakes that will affect the quality of the survey data. Many business survey questionnaires include questions requesting verbal responses, such as those used for classification of the respondent by SIC or type of business, which are subsequently coded by clerks. Most large establishment surveys have survey data entered into a computer by keying, and keyed data are edited in several ways. Records are reviewed for missing or inconsistent data, tabulated survey results are reviewed for possible errors, and data are sometimes imputed by analysts from callbacks to respondents or from other sources of data. Each of these activities provides opportunity for errors. Keying errors, in particular, affect survey results directly and can be very difficult to detect. Coding errors, such as assigning the wrong SIC, will not alter the accuracy of data on individual records, but will cause inaccuracies in survey estimates. Analyst review of tabulations is a subjective

activity at best and errors can occur either by overlooking erroneous results or by overediting results that were correct to begin with. Editing and imputation by analysts are also subjective activities with the same potential problems, with the addition of response errors (caused by interviewer errors) if contacts are made with respondents during editing. Analyst review of data for individual respondents is employed by many of the surveys covered by this report, in contrast to household surveys for which such review is uncommon. This comes from the larger influence on survey results that larger establishments have, thus requiring careful review of data for these larger establishments whereas in a household survey, all households are more or less equally important for survey results, making review by analysts not cost effective for improving data quality.

### **Data Processing by Computer**

Many establishment surveys use computers for much of the processing including editing, imputation, tabulation or computation of estimates, and preparation of survey results for publication. Usually survey requirements are translated into specifications for use in the preparation of computer programs. Both the initial specifications and the resulting programs can alter the original survey plans, thereby leading to errors in individual data records and final results. For example, many surveys use computer programs to perform extensive editing and imputation of individual records. Many ratios, such as payroll to employment, are computed and compared to industry standards. The sheer volume of computations to be programmed suggests that some ratios will be programmed incorrectly or some parameters for these ratios will be built into the programs incorrectly. Even the final tabulations of a census can be programmed incorrectly--for example, aggregating data for the wrong establishments in a publication cell.

### **3. CONTROL OF PROCESSING ERROR**

Various methods are employed in establishment surveys to control the effects of processing errors on survey results. The most common are standard quality control procedures. Acceptance sampling and process control methods are available for such well-defined and easily measured processes as envelope stuffing, clerical coding, and data keying. More subjective processes, such as analyst review of edit failures, do not lend themselves easily to standard quality control methods. However, the processing of surveys is often designed to allow later processing stages to correct errors made in earlier stages. For example, in the processing of the economic censuses, the changes made during the analyst review of failed edit cases are reviewed by sending these cases through the computer edit program that failed the cases originally. While this is not a precise measure of the quality of the analyst review stage, it does serve to limit the errors introduced at this stage of processing.

Two other control procedures are commonly employed to control processing errors in establishment surveys. Interviewers in telephone surveys are usually monitored at least in a supervisory capacity and occasionally in a systematic quality control scheme. This serves to ensure that interviewers follow the prescribed procedures. Also, computer programs are

commonly tested using test files (simulating problems in actual data files) to detect and correct most programming errors. Another technique sometimes used to control computer programming errors is the review of the programming code by the staff that wrote the programming specifications.

#### **4. MEASUREMENT OF PROCESSING ERROR**

##### **Indirect Techniques**

Most large surveys requiring large processing staffs keep performance statistics during processing for supervisory or management purposes. For example, data keying error rates, usually produced from quality control procedures, serve as a supervisory tool with keyers showing high error rates being retrained or fired. Edit failure rates produced during computer editing of survey data provide indications of the expected workload for analysts reviewing the rejected cases. Similarly, the rates of SIC reclassification provide estimates of the workload for other processes. These performance statistics indirectly measure the effects of processing errors on survey data. For the most part, performance statistics provide a count of errors rather than a measure of the effect of errors on data accuracy. For example, quality control procedures can provide an estimate of the percentage of data fields keyed in error, but do not measure the size of the errors included in the total value for a particular data item.

##### **Direct Techniques**

The effect of processing errors on data quality for establishment surveys is rarely measured directly. The opportunity for direct measurement is reduced by the fact that the effects of processing errors are mixed in with response, nonresponse, and coverage errors and cannot be measured separately. For example, in the case of nonresponse errors, it would be impractical to try to measure refusals to respond separately from nonresponse caused by forms mailed to the wrong address. Some special evaluation projects, however, have measured processing errors directly. For example, in the 1982 Economic Censuses, a study was conducted to measure the effect of each processing stage on census data by following the data values for a sample of establishments through the processing. (See U.S. Bureau of the Census, 1987.)

#### **5. SUMMARY PROFILE**

(See Figures 12 and 13.) Standard quality control procedures (process control or acceptance sampling) for data keying and the use of test files for computer programs were the most commonly used controls for the surveys reviewed by this report. This is to be expected since keying is one of the easiest survey operations for which statistical quality control can be used, and the use of test files is common for programming in any context. About half of the surveys used quality control procedures for other activities, including printing, forms checking, coding, and editing. It would be more appropriate for all surveys to use standard quality control procedures for any operations that are repetitive or follow specific guidelines or rules since the use of quality control can greatly reduce errors in these operations. In addition, any clerical operation that can be automated should be, since the opportunity for clerical error is then

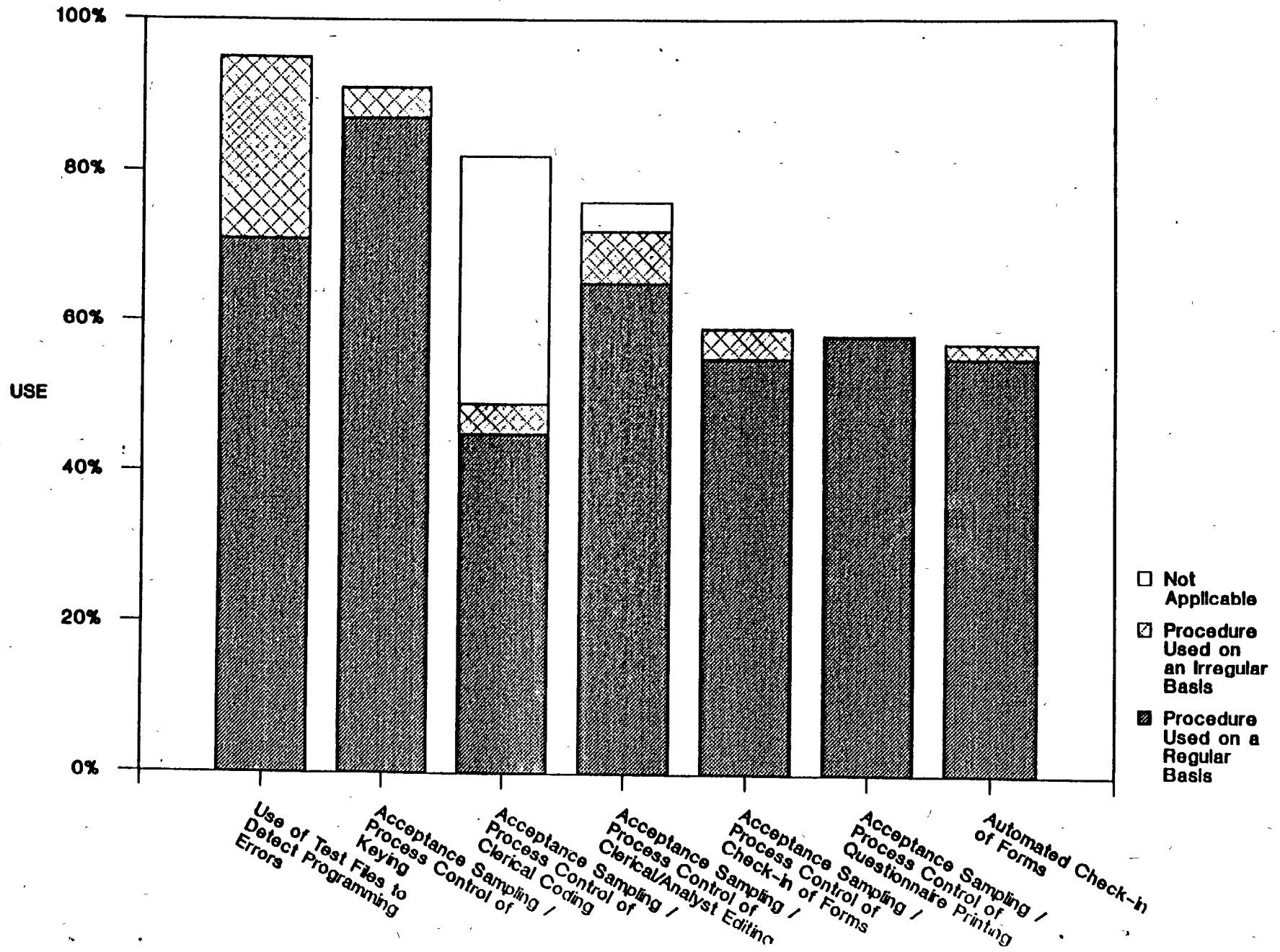
eliminated, such as automated checking of forms used by more than half of the surveys.

About half of the surveys produce keying error rates, edit failure rates and imputation rates which provide indirect measures of processing errors. A few surveys also produce coding error rates and reclassification rates. Almost all of these rates are produced for internal use only however. Some of these rates can be produced as routine output from quality control procedures, so if more surveys employ quality control techniques, more will obtain indirect measures of processing errors. Only one survey reported ever attempting to measure processing errors directly. No indirect measures besides those included in the tables were reported. In summary, survey sponsors and survey takers covered by this report are getting relatively little information about their processing errors.

# PROCESSING ERROR CONTROL PROCEDURES

Extent and Frequency of Use

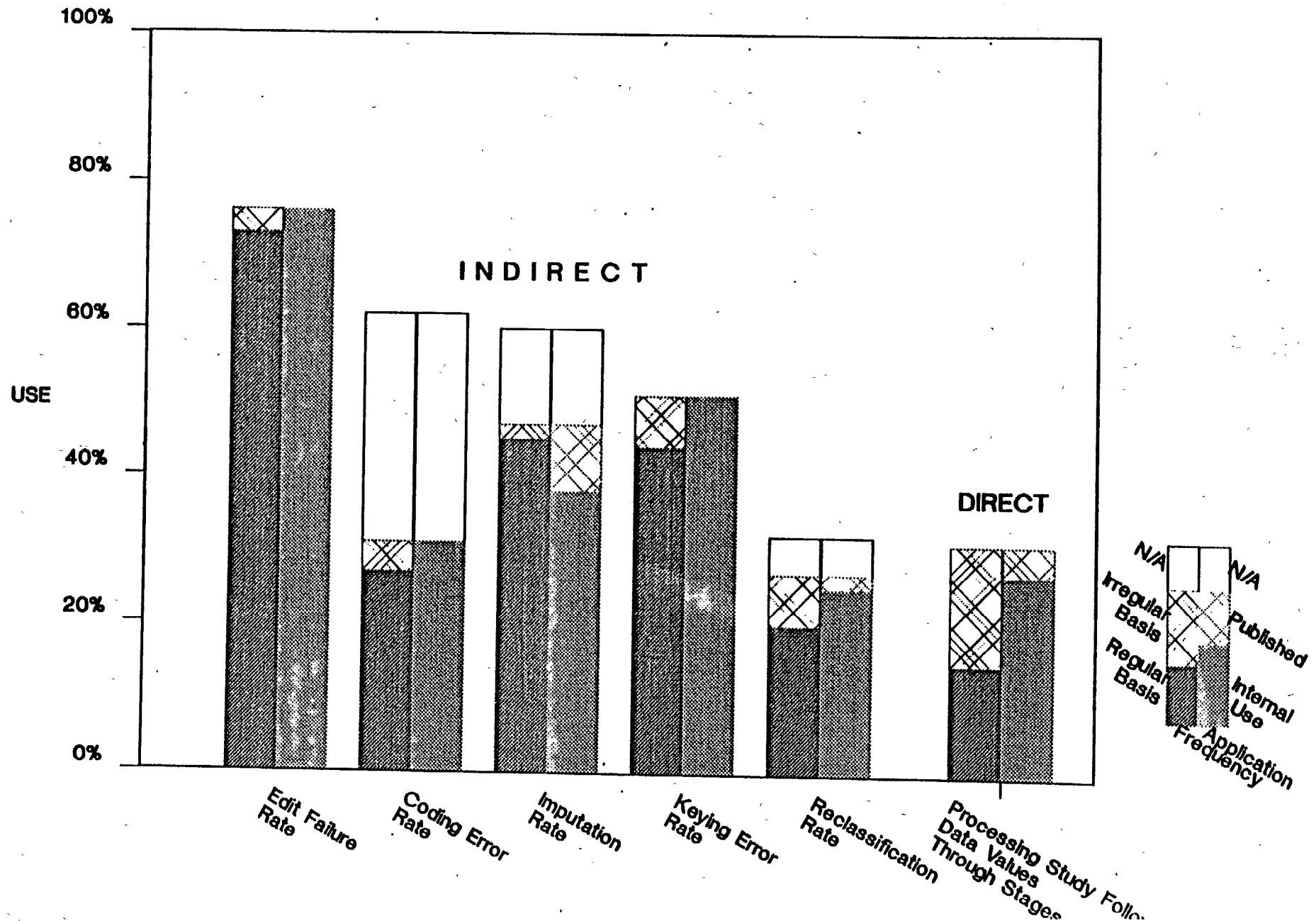
Figure 12



# PROCESSING ERROR MEASUREMENT TECHNIQUES

Frequency and Application of Use

Figure 13



## REFERENCES

- Cochran, William G. 1977. Sampling Techniques. 3d ed. New York: John Wiley and Sons.
- Dalenius, Tore. 1977. "Bibliography of Nonsampling Errors in Surveys." International Statistical Review 45:71-81 and 181-197.
- Fellegi, I.P. 1964. "Response Variance and Its Estimation." JASA 59: 1016-1041.
- Garrett, J., Hogan, H., and Pautler, C. 1986. "Coverage Concepts and Issues in Data Collection and Data Presentation." Proceedings of the Second Annual Research Conference. Washington, D.C.: Bureau of the Census. pp. 329-334.
- Hansen, Morris H., Hurwitz, William N., and Madow, William G. 1953. Sample Survey Methods and Theory, Volume II. New York: John Wiley and Sons.
- Kendall, Maurice G., and Buckland, William R. 1960. 3d ed. A Dictionary of Statistical Terms. Edinburgh: Oliver and Boyd.
- Keyfitz, Nathan. 1951. "Sampling with Probabilities Proportional to Size: Adjustment for Changes in the Probabilities." JASA 46:105-109.
- Kish, Leslie. 1967. Survey Sampling. 2d ed. New York: John Wiley and Sons.
- Konschnik, C., Monsour, N., and Detlefsen, R. 1985. "Constructing and Maintaining Frames and Samples for Business Surveys." Proceedings of the Section on Survey Research Methods, American Statistical Association, 113-122.
- Madow, W. G., Olkin, I., Nisselson, H., and Rubin, D. B., eds. 1983. Incomplete Data in Sample Surveys, three volumes. New York: Academic Press.
- Office of Federal Statistical Policy and Standards. 1978. Glossary of Nonsampling Error Terms: An Illustration of a Semantic Problem in Statistics, Statistical Policy Working Paper 4. Springfield, Va.: National Technical Information Service (PB 86-2115471AS).
- Office of Management and Budget. 1983. Approaches to Developing Questionnaires, Statistical Policy Working Paper 10. Springfield, VA: National Technical Information Service (PB 84-105055).
- Office of Management and Budget. 1987. Standard Industrial Classification Manual. Springfield, VA: National Technical Information Service (PB 87-100012).
- Ostry, Sylvia, and Sunter, Alan. 1970. "Definitional and Design Aspects of the Canadian Job Vacancy Survey." JASA 65:1059-1070.

Rossi, Peter H., Wright, James D., and Anderson, Andy B., eds. 1983. Handbook of Survey Research. New York: Academic Press.

Tenebein, Aaron. 1970. "A Double Sampling Scheme for Estimating from Binomial Data with Misclassification." JASA 65:1350-1461.

U.S. Bureau of the Census. 1987. 1982 Economic Censuses and Census of Governments Evaluation Studies. Washington, D.C.: U.S. Department of Commerce.

U.S. Bureau of the Census. 1986. Proceedings of the Second Annual Research Conference. Reston, VA: U.S. Department of Commerce.

U.S. Bureau of the Census. 1985. Proceedings of the First Annual Research Conference. Washington, D.C.: U.S. Department of Commerce.

U.S. Bureau of the Census. 1974. Standards for Discussion and Presentation of Errors in Data, Technical Paper 32. Washington, D.C.: U.S. Department of Commerce.

United Nations. 1982. National Household Survey Capability Program, Nonsampling Errors in Household Surveys: Sources, Assessment and Control (Preliminary Version), New York: United Nations Department of Technical Cooperation for Development and Statistical Office.

Wolter, Kirk M. 1985. Introduction to Variance Estimation. New York: Springer-Verlag.

Wright, Tommy, ed. 1983. Statistical Methods and the Improvement of Data Quality. Orlando, Fla.: Academic Press.

Zarkovich, S.S. 1966. Quality of Statistical Data. Rome: Food and Agriculture Organization of the United Nations.

## APPENDIX 1

### GOALS, SCOPE, AND USES

#### Goals

- o Document current understanding of what is meant when discussing quality for establishment surveys.
- o Discuss establishment surveys in terms of sampling and nonsampling errors.
- o Profile current practices in the areas of measuring and controlling survey quality.
- o Identify approaches and practices to be considered by users and designers of establishment surveys.
- o Profile major problems in planning, funding, implementing, managing, analyzing, and publishing quality measurement studies.

#### Scope

- o All Federal agencies which conduct establishment surveys will be asked to participate by completing the survey profile collection forms for their surveys.
- o Within an agency, the scope will be limited to all ongoing sample and census surveys. Scope will not cover one time surveys or special studies.
- o All agency major programs should be profiled individually; where a program is made up of numerous small individual surveys (all having similar statistical characteristics), only one composite profile should be developed for the program.
- o The SIC scope will be limited to surveys of the private sector establishments (i.e., exclude strictly government surveys).

#### Uses

- o Establish awareness in sponsors/subject matter specialist of the major error sources associated with establishment surveys.
- o Provide a guide for planning/constructing a basic error profile for an establishment survey.
- o Develop a document which would allow an agency to compare its current survey practices/procedures to other Federal agencies.
- o Use as a framework/standard within an agency to create uniformity of practices.
- o Provide a contrast of establishment error sources vs. household.
- o Basic information document/training document for entry level staff.

## APPENDIX 2

### SURVEY PROFILE QUESTIONNAIRE

#### PROGRAM REQUIREMENTS

1. Is response to the survey voluntary or mandatory?
2. Provide a brief description of the target population for the survey, both in terms of industry (e.g., all industries, service industries, manufacturing industries, hospitals) and geography (e.g., national, 10 largest states, metropolitan areas).
3. What is the level of detail of published tabulations for industry (e.g., two digit SIC only, three digit SIC for manufacturers and two digit SIC for nonmanufacturers) and geography (e.g., national only, national and state)?
4. What are the primary characteristics of interest for the survey (e.g., employment, wages, sales)?
5. What are the estimates of primary interest for the survey? (Mark all that apply.)

<input type="checkbox"/> Level	<input type="checkbox"/> Average	<input type="checkbox"/> Index
<input type="checkbox"/> Change	<input type="checkbox"/> Rate	<input type="checkbox"/> Other-Specify _____

6. What are the design objectives for the survey (e.g., estimate employment for 3-digit industries for the U.S. to within 10,000 at 1 sigma)?
7. What is the frequency of collection and publication?

#### Collection

<input type="checkbox"/> Monthly
<input type="checkbox"/> Quarterly
<input type="checkbox"/> Annually
<input type="checkbox"/> Other-Specify _____

#### Publication

<input type="checkbox"/> Monthly
<input type="checkbox"/> Quarterly
<input type="checkbox"/> Annually
<input type="checkbox"/> Other-Specify _____

8. Where does responsibility for data collection lie?

<input type="checkbox"/> Agency	<input type="checkbox"/> Contractor -	<input type="checkbox"/> Federal
		<input type="checkbox"/> State
		<input type="checkbox"/> Private

## SAMPLE DESIGN

1. Is the survey a sample or census?
2. What is the number of units in the universe?
3. What is the term given for and definition of sampling units?
4. What is the source for the frame (e.g., SSEL, area maps)?
5. What information is available on the frame (e.g., name, address, industry, employment, acreage)?
6. What is the sample size for the survey?
7. Which best describes the survey--repeated with overlap or repeated with no overlap?
8. For surveys with overlap, how long do units remain in sample?
9. Is the survey being implemented as a probability or nonprobability sample?

     Probability

     Nonprobability - Why?

     Substitution is allowed for nonrespondents

     Some large set of units in target population have no chance of selection (e.g., under 10 employees)

     Units are selected judgmentally

     No adequate frame

     Too hard to control

     Other-Specify \_\_\_\_\_

10a. How many stages of sampling are involved?

b. What units are selected at each stage?

c. What sampling technique is used at each stage?

(For Each)

     PPS

     Systematic

     Other-Specify \_\_\_\_\_

11a. What are the primary stratification variables used? (Mark all that apply.)

     Industry

     Sales/Receipts

     Other-Specify \_\_\_\_\_

     Employment

     State

     No stratification

b. What technique is used to allocate sample sizes to strata?

     Proportional

     Neyman

     Optimal

     Other-Specify \_\_\_\_\_

12. Are certainty levels used (i.e., are all firms above a given size selected with certainty)?
13. Is a sample cutoff used (i.e., are all firms below a given size not given a chance of selection)?

## ESTIMATION

1. What types of estimates are published? (Mark all that apply.)

☐ Totals                      ☐ Means                      ☐ Medians  
☐ Proportions              ☐ Indexes                      ☐ Other-Specify \_\_\_\_\_

2. What type of estimation procedure is used?

☐ Expansion/Horvitz-Thompson              ☐ Link relative  
☐ Ratio estimation                      ☐ Other-Specify \_\_\_\_\_

3. Are any independent sources used for adjusting the estimates (e.g., benchmark, ratio)?

4. How are atypical units handled in estimation?

☐ Reweighted                      ☐ No special procedures  
☐ Other-Specify \_\_\_\_\_

5. What type of unit nonresponse adjustment procedure is used?

☐ Weighting Class Adjustment              ☐ None  
☐ Other-Specify \_\_\_\_\_

6. What type of item imputation procedure is used?

☐ Hot Deck                      ☐ Mean Within Class  
☐ Regression                      ☐ Random Within Class  
☐ Other-Specify \_\_\_\_\_ ☐ None

7. Are weighted data used in publication/analysis?

8. What method is used to adjust data for seasonality?

☐ X-11                      ☐ Model based  
☐ Concurrent                      ☐ Other-Specify \_\_\_\_\_  
☐ Projected                      ☐ None

☐ X-11 ARIMA  
☐ Concurrent  
☐ Projected

# VARIANCE ESTIMATION

1. What method of variance estimation is used?

☐ Taylor series  
☐ Balanced repeated  
replication  
☐ Random groups

☐ Generalized variance  
functions  
☐ Other-Specify \_\_\_\_\_  
☐ None

2. Are weighted data used in variance estimation?

3. Is nonresponse adjustment accounted for in variance estimation?

4. Is seasonal adjustment accounted for in variance estimation?

5. At what frequency are variance estimates generated?

☐ Monthly  
☐ Quarterly

☐ Annually  
☐ Other-Specify \_\_\_\_\_

☐ Not generated

6. At what frequency are variance estimates published?

☐ Monthly  
☐ Quarterly

☐ Annually  
☐ Other-Specify \_\_\_\_\_

☐ Not generated

## CONTROL PROCEDURES

(Note: For each, indicates Yes, No, or Not Applicable; if Yes indicate Regular or Irregular Basis.)

### Specification Error

1. Requirements Review--Study to determine what data in a subject area are needed and how they would be used, by contacting potential data users and analysts.
2. Respondent Consultation--Discuss operation definitions, review recordkeeping practices, and explain collection methodology to potential respondents and relevant trade associations, industry representatives, etc.
3. Questionnaire Review by Expert Panel--Convene a panel of experts in the subject matter to review data specifications.
4. Questionnaire Pretest--Study to identify problems such as unclear definitions, poor question wording and instructions, based on analysis of responses given.
5. Cognitive Study--Study to identify problems such as unclear definitions, poor question wording and instructions, based on think-aloud interviews, focus interviews, etc., with respondents.
6. Other--Specify.

### Coverage Error

1. Updating for Structural Changes--Modifying identification information on the frame to account for changes in SIC, company reorganizations, mergers, etc.
2. Integration of Multiple Lists for Frame Development--Use of more than one source of units for development of frame.
3. Sampling from Multiple Frames--Self-explanatory.
4. Use of Independent Control Counts--Use of control counts of units in stratum to check for possible undercoverage/overcoverage.
5. Updating for Sampling of Births--Inclusion (both on the frame and in the sample) of units which come into existence after development of the frame.
6. Internal Consistency Checks for Duplicates--Check of identification information to identify duplicates due to predecessor-successor both listed, partnership listed under all partners' names, same unit contained on several lists, etc.

7. Conduct Special Frame Improvement Surveys--Survey conducted to improve coverage of frame.
8. Enlarge Scope to Capture All Units of Interest--Spread target population to aid coverage of units of actual interest.
9. Internal Consistency Checks for Frame Content--Check of identification information to identify incorrect information.
10. Use of Two-Stage Sampling--Use of multiple stages of sampling to avoid errors due to incorrect information on frame.
11. Include as In-Scope Units with Changed Address, Geography, Industry, Size--Retain units with errors in AGIS, rather than drop them as out of scope.
12. Include Units Closed for Season--Retain units temporarily closed.
13. Sample Validation--Comparison of sample units weighted up to universe totals for certain characteristics.
14. Obtain Information from Units to Allow Reweighting--Obtain identification information when unit of interest cannot be adequately identified in field; for use in reweighting.
15. Followup for Inadequate Mail Address--Self-explanatory.
16. Other--Specify.

#### Response Error

1. Recordkeeping Practices Study--Study to investigate respondent's recordkeeping practices to design collection of items consistent with those practices.
2. Questionnaire Pretest--Study to identify potentials for response error based on analysis of responses given.
3. Cognitive Study--Study to identify potentials for response error, based on think-aloud interviews, focus interviews, etc., with respondents.
4. Use of Administrative Data in Editing--Use of administrative data on sample units to identify and/or impute for potential response errors.
5. Analyst Review of Data--Use of subject matter specialist to identify and/or impute for potential response errors.
6. Personal Visit Initiation--Self-explanatory.
7. Edit Data for Reasonableness and Develop Followup Procedures--Self-explanatory.

8. Reinterview Sample of Interviewer's Work--Self-explanatory.
9. Detailed Training/Guidelines for Interviewers--Self-explanatory.
10. Use of CATI with On-line Monitoring--Self-explanatory.
11. Records Check Study--Study reviewing respondent's hard data to identify response error.
12. Eliminate Nonmeasurable Items--Self-explanatory.
13. Other--Specify.

#### Nonresponse Error

1. Process Control/Acceptance Sampling of Mailing Operation--Verify forms sent to all sample units.
2. Central Office Clearance and Special Reporting Arrangements--Arrangements made to assist reporting for large units.
3. Intensive Followup of Critical Units--Extra effort put into obtaining responses from large units.
4. Provide Survey Publication to Sample Units--Self-explanatory.
5. Use of Overlap Techniques to Maintain Large Units in Sample--Self-explanatory.
6. Advance Notification--Self-explanatory.
7. Use of Unit Nonresponse Followup--Self-explanatory.
8. Rotate Sample for Small Units--Self-explanatory.
9. Personal Visit Initiation--Self-explanatory.
10. Data-Keeping Practices Pilot Test--Investigate respondent's recordkeeping practices to design collection minimize nonresponse.
11. Use of Item Nonresponse Followup--Self-explanatory.
12. Other--Specify.

#### Processing Error

1. Process Control/Acceptance Sampling of Check-In of Forms--Self-explanatory.
2. Automated Check-In of Forms--Self-explanatory.

3. Process Control/Acceptance Sampling of Clerical Coding--Self-explanatory.
4. Process Control/Acceptance Sampling of Keying--Self-explanatory.
5. Process Control/Acceptance Sampling of Clerical/Analyst Editing--Self-explanatory.
6. Use of Test Files to Detect Programming Errors--Self-explanatory.
7. Other--Specify.

## MEASUREMENT TECHNIQUES

(Note: For each indicate Yes, No, or Not Applicable; if Yes indicate Regular or Irregular Basis, and whether Published or Internal Use Only. If a procedure is listed, question refers to whether appropriate measures for analysis are produced.)

### Specification Error

#### Indirect Measures

1. Questionnaire Pretest
2. Cognitive Study
3. Comparison to Independent Estimates
4. Other

#### Direct Measures

1. Records Check Study
2. Other

### Coverage Error

#### Indirect Measures

1. Birth Rate
2. Out of Business Rate
3. Out of Scope Rate
4. Unclassified Rate
5. Misclassification Rate
6. Duplication Rate
7. Sample Attrition Rate
8. Evaluation of OOS/Nonexistent Classifications
9. Other

#### Direct Measures

1. Post Enumeration Survey
2. Comparison of Characteristics of Target and Covered Populations Based on Independent Estimates
3. Match Known Population Units Against Frame Units
4. Check Frame Against Alternative Lists
5. Recheck of Interviewers' Unit Listings
6. Studies of Components of the Frame
7. Other

## Response Error

### Indirect Measures

1. Edit Failure Rate
2. Interviewer Error Rate
3. Questionnaire Pretest
4. Cognitive Study
5. Other

### Direct Measures

1. Records Check Study
2. Detailed Content/Reinterview Study
3. Interviewer Variance Study
4. Response Variance Study
5. Other

## Nonresponse Error

### Indirect Measures

1. Unit Response Rate
2. Weighted Unit Response Rate
3. Item Response Rate
4. Item Coverage Rate
5. Refusal Rate
6. Distribution of Reason for Nonresponse
7. Comparison of Data Across Contacts
8. Other

### Direct Measures

1. Link to Administrative Data for Nonrespondents
2. Followup Sample of Nonrespondents to Estimate Nonresponse Bias
3. Other

## Processing Error

### Indirect Measures

1. Keying Error Rate
2. Coding Error Rate
3. Edit Failure Rate
4. Imputation Rate
5. Reclassification Rate
5. Other

### Direct Measures

1. Processing Study Following Data Values Through Processing Stages
2. Other

## APPENDIX 3

### PROFILE OF SURVEY PRACTICES

#### Federal Establishment Surveys Covered

##### Bureau of Labor Statistics (13 programs)

###### Current Employment Statistics Survey

Monthly survey of 280,000 establishments collecting information on total employment, women and production workers, hours, and earnings

###### Occupational Employment Statistics Survey

Annual survey collecting employment by occupation from sample of approximately 600,000 establishments, using a three-year cycle

###### Hours at Work Survey

Annual survey of approximately 4,600 establishments collecting information on hours worked and hours paid

###### Occupational Safety and Health Survey

Annual survey of approximately 300,000 establishments collecting data on occupational illnesses and injuries

###### Area Wage Surveys

Set of 71 annual surveys collecting local area wages by occupation by industry division

###### Industry Wage Surveys

Set of surveys on selected industries run on a roughly 5-year cycle collecting information on earnings by occupation for the industry

###### Professional, Administrative, Technical, and Clerical Survey

Annual survey of approximately 5,500 establishments collecting data on wages by occupation

###### Service Contract Area Surveys

Set of 92 annual surveys collecting information on earnings by occupation

###### Employee Benefit Survey

Annual survey of 1,500 establishments collecting data on employee benefits

###### Employment Cost Index Survey

Quarterly survey collecting wages and benefits information from a sample of 3,000 establishments

###### Consumer Price Index - Commodities and Services

Monthly survey of approximately 35,000 retail establishments collecting price information on commodities and services

**International Producer Price Survey**

Quarterly survey collecting information on prices for selected product groups from approximately 5,600 establishments with imports into U.S. or exports out of U.S.

**Producer Price Index Survey**

Monthly survey of approximately 40,000 mining and manufacturing establishments collecting information on selected product groups

**Census Bureau (15 programs)**

**Census of Agriculture**

Quinquennial census of all farm operators collecting data on livestock and crop quantities, operator characteristics, and expenditures

**Census of Construction Industries**

Quinquennial survey of 180,000 construction establishments collecting data on employment, payroll, and receipts

**Quarterly Financial Report for Manufacturing, Mining, and Trade Corporations**

Quarterly survey of 306,000 enterprises in wholesale, retail, mining, and manufacturing, collecting income and balance sheet information

**Censuses of Retail, Wholesale, Trade, and Services (3 programs)**

Quinquennial census of: about 1.3 million employer establishments in retail trade; about 415,000 employer establishments in wholesale trade; and about 1.3 million employer establishments in selected service industries; collects information on sales/receipts, employment, payroll

**Current Surveys of Retail, Wholesale, Trade, and Services (3 programs)**

Monthly surveys of: retail sales and inventory of about 11,000 firms and 48,000 establishments for sales and about 3,900 firms for inventory; merchant wholesale sales and inventory of about 3,200 firms. Annual surveys in retail trade (about 28,000 firms), merchant wholesale (about 7,000 firms), and selected services (about 27,000 firms); collects information on sales/receipts, inventory, purchases

**Annual Survey of Manufactures**

Survey of about 55,000 manufacturing establishments collecting data on value of shipment and product class shipment

**Current Industrial Reports**

Collection of monthly, quarterly, and annual surveys of manufacturing industries collecting data on shipments and production

**Survey of Industrial Research and Development**

Annual survey of 12,000 companies in manufacturing and selected nonmanufacturing industries collecting information on expenditures for research and development

**Manufactures Shipments, Inventories, and Orders Survey**

Monthly survey of 4,100 manufacturing plants collecting information on shipments, orders, and inventories

**Survey of Pollution Abatement Costs and Expenditures**

Annual survey of 20,000 manufacturing establishments collecting information on total expenditures made to abate pollution emissions

**Survey of Plant Capacity**

Annual survey of 9,000 manufacturing establishments measuring preferred and practical levels of capacity utilization

**Energy Information Administration (9 programs)**

**Coal Distribution Report**

Quarterly census of coal mines and companies, collecting information on origin of coal, distribution, sales, and stocks

**Coal Production Report**

Quarterly census of coal mines and companies collecting information on production, disposition, and productivity

**Monthly Power Plant Report**

Monthly census of electric utilities collecting information on net generation, fuel consumption, and fuel stocks

**Monthly Report of Cost and Quality of Fuels for Electric Power Plants**

Monthly census of electric utilities collecting information relative to sale of fuels

**Monthly Report of Natural Gas Purchases and Deliveries to Consumers**

Survey of 390 natural gas companies collecting information on volumes and prices

**Annual Report of Natural Gas and Supplemental Gas Supply and Disposition**

Survey of natural and synthetic gas producers, processors, distributors, and pipelines collecting information on origin of supplies, disposition, and price

**Monthly Petroleum Product Sales Report**

Survey of 2,700 refiners and gas plant operators collecting information on sales prices and volumes of selected petroleum products

**Reseller/Retailer's Monthly Petroleum Product Sales Report**

Survey of distillate fuel oil resellers/retailers, motor gasoline wholesalers and residual fuel oil resellers/retailers collecting information on sales volumes and prices

**Petroleum Supply Surveys**

Set of weekly, monthly, and annual surveys of petroleum refineries and blending plants, bulk terminals, product pipeline companies, and importers collecting information on production, imports, and stocks of petroleum

Federal Reserve Bank (6 programs)

Consumer Installment Credit

Monthly survey of components of consumer lending at 400 insured commercial banks

Debits to Demand and Savings Deposits Accounts

Monthly survey of debits to (withdrawals from) selected accounts at 300 insured commercial banks

Report of Transaction Accounts, Other Deposits, and Vault Cash

Weekly survey of 12,000 financial institutions collecting levels of money stock deposit items

Terms of Bank Lending

Quarterly survey of 348 insured commercial banks collecting information on business and farm loans

Monthly Survey of Selected Deposits

Collects data on deposit levels and interest rates paid at 600 insured commercial and savings banks

Weekly Report of Selected Assets

Collects levels of asset items from 1,100 insured commercial banks

National Agricultural Statistics Service (4 programs)

Farm Costs and Returns Survey

Annual survey of 24,000 farms collecting data on financial conditions, production expenses, capital expenditures, and production practices

June Enumerative Survey

Annual survey of 16,000 land segments collecting data on livestock, planted crops, and grain stocks

Objective Yield Survey

Monthly surveys (in season) of farm fields collecting crop production data for eight crops

Quarterly Agricultural Survey

Survey of 80,000 farms collecting data on acreage and production of crops, grain stocks and capacity, and livestock totals

Center for Education Statistics (3 programs)

National Survey of Private Schools

Biennial survey of 1,700 private elementary and secondary schools collecting information on school and teacher characteristics

Survey of Public and Private School Libraries and Media Centers

Quinquennial survey of 6,200 elementary and secondary schools collecting information on library collections, services, staffing, and expenditures

**Post Secondary Education Surveys**

Set of annual and biennial surveys of 6,200 public and private post-secondary institutions collecting information on enrollment, programs, expenditures, revenues, salaries, etc.

**Bureau of Economic Analysis (2 programs)**

**Plant and Equipment Survey**

Quarterly (12,000 units) and annual (9,000 units) survey of nonagricultural enterprises collecting information on current and planned plant and equipment expenditures

**Surveys of Foreign Affiliated Businesses**

Quarterly, annual, and quinquennial surveys of business enterprises with foreign affiliation (either partial ownership of or partially owned by foreign business) collecting information on transactions, financial and operational data, balance sheets, income statements

**Bureau of Mines (2 programs)**

**Iron and Steel Scrap Survey**

Monthly survey of 400 establishments consuming iron and steel scrap collecting information on receipts, production, consumption, shipments, and stocks

**Ferrous/Nonferrous/Industrial Minerals Surveys**

Set of monthly, quarterly, and annual censuses of establishments consuming, producing, or shipping minerals collecting information on receipts, production, consumption, shipments, and stocks

**National Center for Health Statistics (1 program)**

**National Nursing Home Survey**

Periodic (generally every 4 years) survey of 1,200 nursing homes collecting information on occupancy, discharges, and resident characteristics

**Reports Available in the  
Statistical Policy  
Working Paper Series**

1. Report on Statistics for Allocation of Funds (Available through NTIS Document Sales, PB86-211521/AS)
2. Report on Statistical Disclosure and Disclosure-Avoidance Techniques (Available through NTIS Document Sales, PB86-211539/AS)
3. An Error Profile: Employment as Measured by the Current Population Survey (Available through NTIS Document Sales PB86-214269/AS)
4. Glossary of Nonsampling Error Terms: An Illustration of a Semantic Problem in Statistics (Available through NTIS Document Sales, PB86-211547/AS)
5. Report on Exact and Statistical Matching Techniques (Available through NTIS Document Sales, PB86-215829/AS)
6. Report on Statistical Uses of Administrative Records (Available through NTIS Document Sales, PB86-214285/AS)
7. An Interagency Review of Time-Series Revision Policies (Available through NTIS Document Sales, PB86-232451/AS)
8. Statistical Interagency Agreements (Available through NTIS Document Sales, PB86-230570/AS)
9. Contracting for Surveys (Available through NTIS Document Sales, PB83-233148)
10. Approaches to Developing Questionnaires (Available through NTIS Document Sales, PB84-105055/AS)
11. A Review of Industry Coding Systems (Available through NTIS Document Sales, PB84-135276)
12. The Role of Telephone Data Collection in Federal Statistics (Available through NTIS Document Sales, PB85-105971)
13. Federal Longitudinal Surveys (Available through NTIS Document Sales, PB86-139730)
14. Workshop on Statistical Uses of Microcomputers in Federal Agencies (Available through NTIS Document Sales, PB87-166393)
15. Quality in Establishment Surveys (Available through NTIS Document Sales, PB88-232921)

Copies of these working papers may be ordered from NTIS Document Sales, 5285 Port Royal Road, Springfield, VA 22161 (703) 487-4650